Academic Year 2015-2016

Ghent University-Faculty of Sciences

Flanders Institute for Biotechnology

Inflammation Research Center

Medical Biotechnology Center

IDENTIFICATION AND FUNCTIONAL ANALYSIS OF MOLECULAR INTERACTION PARTNERS OF δ–PROTOCADHERIN FAMILY MEMBERS WITH PUTATIVE ROLES IN CANCER

Eleonora Billi

Thesis submitted in partially fulfilment of the requirements for the degree of

Doctor of Science: Biochemistry and Biotechnology

Promoters:

Prof. Dr. Frans van Roy and Prof. Dr. Jan Tavernier

Identification and functional analysis of molecular interaction partners of δ–Protocadherin family members with putative roles in cancer

Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor in Science: Biochemistry and Biotechnology

By Eleonora Billi

Molecular Cell biology Unit

Inflammation Research Center (IRC), Ghent University

Technologiepark 927

B-9052 Ghent-Zwijnaarde

Cytochine Receptor Lab (CRL)

Medical Biotechnology Center (MBC), Ghent University

Albert Baertsoenkaai 3

B-9000 Ghent

Promoters: *Prof. Dr. Frans van Roy (Ghent University, Belgium)*

Prof. Dr. Jan Tavernier (Ghent University)

Examination committee:

Chairman*: Prof. Dr. Johan Grooten (Ghent University, Belgium)*

Secretary*: Prof. Dr. Geert Berx (Ghent University, Belgium)*

Members*: Prof. Dr. Jean Christophe Marine (KU Leuven, Belgium)*

Prof. Dr. Marleen van Troys (Ghent University, Belgium)

Dr. Irma Lemmens (Ghent University, Belgium)

Dr. Karl Vandepoele (Ghent University Hospital, Belgium)

To Alessandro

Il corpo faccia quello che vuole. Io non sono il corpo: io sono la mente.

The body does whatever it wants. I am not my body: I am my mind.

Rita Levi Montalcini, 1909-2012, italian neurologist.

So Mary, climb in.

Table of contents

List of acronyms

aa	amino acid
AD	Activation Domain
AP	Affinity Purification
ASD	Autism Spectrum Disorders
BHR	Bronchial Hyper responsiveness
BiFC	Biomolecular Fluorescence Complementation Assay
BRET	Bioluminescence Resonance Energy Transfer
C-Pcdh	Clustered Protocadherin
CAT	Chloramphenicol Acetyltransferase
CCC	Cadherin-Catenin-Complex
CD	Cytoplasmic Domain
CDHR5	Cadherin-Related Family Member 5
CLL	Chronic Lymphocytic Leukemia
CM	Conserved Motif
CNS	Central Nervous System
CoIP	Coimmunoprecipitation
DBD	DNA Binding Domain
DMEM	Dulbecco's Modified Eagle's Medium
DOX	Doxycycline
DUB	Deubiquitinating Enzymes
DYNLT1	Dynein Light Chain, Tctex-Type 1
EC	Extracellular Cadherin Repeat
EGL	External Granule Cell Layer
EMFR	Epilepsy And Mental Retardation Limited To Females
EMT	Epithelial-To-Mesenchymal Transition
EpoR	Erythropoietin Receptor
ES cell	Embryonic Stem Cells
EYFP	Enhanced Yellow Fluorescent Protein
FC	Fluorescence Complementation
FHL	Four And A Half Lim Domains
FPKM	Fragments Per Kilo base Of Transcript Per Million Mapped Reads
FRET	Förster Resonance Energy Transfer
FS	Febrile Seizures
GFP	Green Fluorescent Protein
GHR	Growth Hormone Receptor
GO	Gene Ontology
GST	Glutathione S-Transferase
HRP	Horseradish Peroxidase
HSP	Heat Shock Protein
IC	Intermediate Chain
IF	Immunofluorescence

IP	Immunoprecipitation
JAK	Janus Kinase
kb	kilo base
KD	Knock Down
KISS	Kinase Substrate Sensor
KO	Knock Out
LC	Light Chain
LIC	Light Intermediate Chain
LR	Leptin Receptor
MACS	Magnetic Cell Sorting
MaMTH	Mammalian-Membrane Two-Hybrid Assay
MAPPIT	Mammalian Protein-Protein Interaction Trap
MASPIT	Mammalian Small Molecule-Protein Interaction Trap
MAX	Myc Associated Factor X
MeCP2	Methyl-Cpg-Binding Protein 2
MEF2	Myocyte Enhancer Factor-2
MS	Mass Spectrometry
MTX	Methotrexate
MXD3/MAD3	Max Dimerization Protein 3
MY2H	Membrane Yeast Two-Hybrid
Nap1	Nck-Associated Protein 1
NBF	Nap1- Binding Fragment
NC- Pcdh	Non Clustered Protocadherin
NLS	Nuclear Localization Signal
NSCLC	Non-Small-Cell Lung Cancer
ORF	Open Reading Frame
PAPC	Paraxial Protocadherin
PBS	Phosphate Buffer Saline
PCA	Protein Fragment Complementation Assay
PCDH	Protocadherin
PCDH19-FE	Pcdh19 Related Female Epilepsy
PDLIM7/Enigma	Pdz-Lim Family Member 7
PEG	Polyethylene Glycol
PLA	Proximity-Ligation Assay
PMSC1	Proteasome 26S Subunit, ATPase, 1
PP1α	Protein Phosphatase 1A
PPI	Protein-Protein Interaction
PSD	Post Synaptic Density
PYK2	Proline-Rich Tyrosine Kinase 2
qRT-PCR	Reverse Transcription Quantitative Real-Time Pcr
RB	Retinoblastoma
RET	Resonance Energy Transfer
RET	Receptor Tyrosine-Protein Kinase
rPAP	Rat Pancreatitis-Associated Protein I

rRRS	Reverse Ras Recruitment System Yeast Two Hybrid
RRS Y2H	Ras Recruitment System Yeast 2 Hybrid
RT	Room Temperature
RTA system	Trans-Activator System
SH	Src Homology Domains
shh	Sonic Hedgehog
SNP	Single Nucleotide Polymorphism
STAT	Signal Transducer And Activator Of Transcription
SZ	Schizophrenia
TAP	Tandem Affinity Purification
TLR	Toll Like Receptor
TM	Transmembrane Domain
TSG	Tumor Suppressor Gene
Ub	Ubiquitin
UPP	Ubiquitin-Proteasome Pathway
USP	Ubiquitin Specific Peptidase
WASP	Wiskott–Aldrich Syndrome Protein
WAVE	Wasp-Family Verprolin-Homologous Protein
WIRS	WRC-Interacting Receptor Sequence
WRC	Wave Regulatory Complex
Y2H	Yeast Two-Hybrid
Y3H	Yeast Three-Hybrid
YFP	Yellow Fluorescent Protein

Summary

The subject of this dissertation is the transmembrane protein family Delta-Protocadherin (δPCDH), belonging to the superfamily of cadherin. δPCDH structure shows an extracellular domain related to E-cadherin and a unique intracellular domain. The functions of the members of the family are still unclear. The family includes 10 members further divided in δ0PCDH, δ1PCDH and δ2PCDH subgroups depending on their extracellular cadherin domains number and the presence (or absence, for Pcdh20) of 2 or 3 conserved motifs (CM) in their cytoplasmic tail. Studies have shown that PCDHs are probably rather communicative than adhesive proteins and functional differences may exist between long and short isoforms. PCDHs have been found mutated in several cancer types and neurodegenerative diseases. In recent years, direct and indirect evidences have accumulated indicating an important role for δPCDHs during tumorigenesis and tumor suppression. Promoter methylation and transcriptional silencing of *δPCDH* genes have been shown to occur in numerous epithelial cancer types and in multiple hematological malignancies, gliomas and medulloblastomas. Moreover, loss of expression of various δPcdhs (PCDH1, PCDH9, PCDH10, PCDH17 and PCDH20) often correlates with poor prognosis or with therapy resistance. For some PCDHs, such as PCDH11Y, an oncogenic role has been indicated: PCDH11Y expression correlates with the hormone-independent growth of progressed prostate tumors. Also PCDH7 expression has been linked to tumor progression rather than to suppression; *PCDH7* is an upregulated gene in brain metastasis of breast cancer models. PCDHs are expressed in high levels in brain so it is not surprising that many of them are involved in neurological disorders such as autism, schizophrenia, epilepsy and Alzheimer's diseases. A role in mental diseases has been shown for PCDH7, PCDH10, PCDH8, PCDH17, PCDH11X and PCDH11Y, and PCDH18, but the most convincing evidence comes from the correlation of PCDH19 and epilepsy. PCDHs have been related to other kinds of disease as well, including asthma and bronchial hyperresponsiveness.

To investigate the functions of PCDHs, we performed studies of protein-protein interaction, in order to identify novel interaction partners of the members of the family. A screening of a collection of 10000 ORFs was performed with Array mammalian protein-protein interaction trap (MAPPIT) technology for PCDH9, PCDH10 and PCDH11X. One-to-one binary MAPPIT assay and Ingenuity Pathway Analysis (IPA) were performed to obtain an experimental/literature-based priority list of candidate interaction partners for different members of the family. Co-Immunoprecipitation

experiments were performed as a biochemical confirmation of the interactions. Investigation of the functional relevance of those interactions was not straightforward and still requires additional studies. RNA expression levels of identified putative interaction partners and related proteins were investigated with qRT-PCR. The tests were performed in RNA extracted from cerebella or total brain of wild type (WT) and knock out (KO) mice for Pcdh10. For our experiments, we generated two different kinds of KO mice for Pcdh10: the first showing only the short isoforms of the protein (Pcdh10long KO), while in the second KO mouse all the isoforms were deleted (Pcdh10all KO). RNA from cerebellum has been used as well to perform RNA sequencing experiments: surprisingly the Pcdh10all KO mouse did not show change in RNA expression levels. Pcdh10long KO mouse for the long isoforms only shows differences in RNA expression levels for several genes if compared with the WT. We speculate that in Pcdh10all KO mice, compensation events from the other members of the family might occur while in Pcdh10long KO mice the short isoform inhibits the compensation mechanisms of the other PCDHs. In support of this we performed qRT-PCR analysis of both $\delta 1$ and $\delta 2$ Pcdh family members in brain of Pcdh10all KO mice and it revealed an upregulation of Pcdh7, Pcdh8, Pcdh11, Pcdh17 and Pcdh18 transcript while none of the δPcdhs showed a change in the Pcdh10long KO mice. Results of RNA sequencing were analyzed via IPA and Metacore Analysis.

Beside expression level investigations, studies of ubiquitination were performed, since a lot of candidate belongs to the ubiquitin family. We could show that PCDH10, PCDH11X, PCDH1 and PCDH7 can be ubiquitinated and that the novel interaction partners do not play a role in this mechanism. We also performed tests of protein stability to assess if it would change in presence of ubiquitin-proteasome system (UPS) proteins: we could define a range of turnover of about 20-36 hours but we could not show difference in presence of UPSs. To check the relevance of another family of candidates we performed MYC-related tests such as MYC reporter assay and binary MAPPIT with doxorubicin to induce the p53 pathway but we could not show difference linked to the interaction with PCDHs. We then performed a quantitative analysis of cell properties in cell lines with induced PCDH expression: experiments on migration, proliferation and adhesion were carried out to investigate if PCDHs play a physiological role in these processes. Our preliminary results are promising: we established in *vitro* cell-based assay to evaluate the effect of PCDH10 or PCDH11X expression and this can be used in subsequent assays to test the relevance of the molecular partner interactions that we identified. Moreover, via the MAPPIT array we identified interesting proteins as putative interaction partners of PCDHs. In this dissertation we attempt to elucidate the relevance of these interactions aiming to identify the mechanisms of PCDHs in health and in pathological conditions.

Samenvatting

Het onderwerp van deze doctoraatsthesis is de transmembraaneiwitfamilie Delta-Protocadherine (δPCDH), die behoort tot de cadherine-superfamilie. δPCDHs worden gekenmerkt door een uniek intracellulair domein en een extracellulair domein dat gerelateerd is tot E-cadherine. Daarnaast is de functie van de leden van deze eiwitfamilie tot op heden onbekend. De δPCDH-familie bevat 11 leden. Deze zijn verder onderverdeeld in de subgroepen δ0PCDH, δ1PCDH en δ2PCDH op basis van hun extracellulaire cadherine-domeinnummers en de aanwezigheid (of in het geval val Pcdh20 de afwezigheid) van twee of drie geconserveerde motieven (CM) in hun cytoplasmatische staart. Eerdere studies hebben aangetoond dat PCDHs eerder een communicatieve dan adhesieve eiwitfunctie hebben en dat er functionele verschillen bestaan tussen de lange en korte isovormen. Recent is zowel direct als indirect bewijs aangeleverd dat een belangrijke rol suggereert voor de δPCDHs tijdens tumorigenese en tumorsuppressie. Daarbij is aangetoond dat promotormethylatie en transcriptiesilencing van de δPCDH genen in verschillende epitheliale kankertypes en hematologische maligniteiten, gliomen en medulloblastomen voorkomen. Bovendien is een expressieverlies van verschillende δPcdhs (PCDH1, PCDH9, PCDH10, PCDH17 en PCDH20) vaak gecorreleerd met therapieresistentie of een slechte prognose. Voor sommige PCDHs wordt er tevens een oncogene rol gesuggereerd. Zo is de expressie van PCDH11Y gecorreleerd met de hormoononafhankelijke groei van gevorderde prostaattumoren. Ook de expressie van PCDH7 wordt eerder gelinkt aan tumorprogressie dan suppressie, aangezien PCDH7 upgereguleerd is tijdens de metastasis in de hersenen en in borstkankermodellen. In de hersenen komt tevens een sterke expressie van PCDHs voor, waardoor het niet verrassend is dat vele PCDHs eveneens gecorreleerd zijn met neurologische storingen zoals autisme, schizofrenie, epilepsie en de ziekte van Alzheimer. Hierbij wordt er een belangrijke rol gesuggereerd voor PCDH7, PCDH10, PCDH8, PCDH17, PCDH11X and PCDH11Y en PCDH18; hoewel het sterkste bewijs is aangetoond voor de correlatie tussen PCDH19 en epilepsie. Verder zijn PCDHs ook in verband gebracht met andere ziektes zoals astma en bronchiale hyperreactiviteit.

In deze doctoraatsthesis zijn de functies van PCDHs verder onderzocht. Meer specifiek zijn er nieuwe interactiepartners van de PCDH-familieleden geïdentificeerd met behulp van eiwit-eiwit interactieëxperimenten. Er werd een screening uitgevoerd op een collectie van 10000 ORFs met behulp van een Array mammalian protein-protein interaction trap (MAPPIT) voor PCDH9, PCDH10 en PCDH11X. Vervolgens werd er een prioriteitslijst opgesteld van potentiële interactiepartners voor de

verschillende eden van de PCDH-familie. Hiervoor werd gebruik gemaakt van één-op-één binaire MAPPIT assays en Ingenuity Pathway Analysis (IPA). Tevens werden er co-immunoprecipitatietesten uitgevoerd met als doel een biochemische confirmatie van de geïdentificeerde interacties te bekomen. Het bestuderen van de functionele relevantie van deze interacties is evenwel niet voor de hand liggend en vergt verder onderzoek.

Vervolgens werden de RNA expressieniveaus van de geïdentificeerde vermeende interactiepartners en de gerelateerde eiwitten onderzocht aan de hand van qRT-PCR. Deze tests werden uitgevoerd met RNA geëxtraheerd van de cerebella of van de gehele hersenen van wild type (WT) en knock out (KO) muizen voor Pcdh10. Voor deze experimenten werden twee types van KO muizen voor PCDH10 gebruikt. Het eerste type miste enkel de lange isovorm van het eiwit (Pcdh10long KO), terwijl bij het tweede type KO muis alle isovormen gedeletet werden (Pcdh10all KO). RNA van het cerebellum werd eveneens gebruikt voor RNA sequencing. Opmerkelijk werd hierbij aangetoond dat de Pcdh10all KO muizen geen verschil in RNA expressieniveaus vertoonden, terwijl voor de Pcdh10long KO muizen een lange lijst van up- en down-regulatie van genexpressie werd bekomen. Er werd daarom gespeculeerd dat in de Pcdh10all KO muizen compensatie-events plaatsvinden door andere leden van de PCDH-familie, terwijl voor de Pcdh10long KO muizen de korte isoform deze compensatiemechanismen verhindert. Deze hypothese werd bevestigd door de qRT-PCR analyses van zowel $\delta 1$ en $\delta 2$ Pcdh familieleden van de muizenhersenen. In het geval van Pcdh10all KO muizen werd een up-regulatie aangetoond van Pcdh7, Pcdh8, Pcdh11, Pcdh17 en Pcdh18 transcripts, terwijl de δPcdhs geen verandering vertoonden in de Pcdh10long KO muizen. De resultaten van de RNA sequencing werden overigens geanalyseerd via IPA en Metacore Analyses.

Aangezien vele van de geïdentificeerde kandidaten tot de ubiquitine familie behoren, werd naast de expressieniveaus ook de ubiquitinatie bestudeerd. Meer specifiek werden daarbij experimenten uitgevoerd in Hek293T cellen tijdens transiënte transfectie. Eerst werd er gecontroleerd of de PCDHs geübiquitineerd waren met behulp van PCDH10, PCDH11X, PCDH1 and PCDH7. Zodra er ubiquitinatie van de PCDHs geobserveerd werd, werden de cellen gecotransfecteerd met een vermeende ubiquitine-gerelateerde interactiepartner. Deze experimenten konden evenwel geen verandering aantonen in de ubiquitinatieniveaus in de aanwezigheid van de geselecteerde kandidaten. Verder werd er gepostuleerd dat de ubiquitine-proteasome systeem (UPS) eiwitten eveneens kandidaat-interactiepartners zijn. Om deze reden werden er proteïnestabiliteitstesten uitgevoerd om te achterhalen of de proteïnestabiliteit zou wijzigen in de aanwezigheid van deze UPS eiwitten. Hoewel er een omzettingsrange van 20 tot 36 uur werd bekomen, konden er geen

verschillen worden vastgesteld tussen de condities met aan- en afwezigheid van de UPS. Om de relevantie van ook een andere familie van kandidaten te kunnen onderzoeken, werden eveneens MYC-gerelateerde testen uitgevoerd zoals MYC reporter assays en binaire MAPPIT met doxorubicine om de p53 pathway te induceren. Hierbij werden echter geen verschillen waargenomen die gelinkt zijn met een interactie met PCDHs. Vervolgens werd een kwantitatieve analyse van de celeigenschappen in de cellijnen uitgevoerd met behulp van geïnduceerde protocadherine expressie. Meer specifiek werden er experimenten uitgevoerd met betrekking tot migratie, proliferatie en adhesie, om op deze manier te kunnen onderzoeken of PCDHs een fysiologische rol spelen in deze processen. De bekomen preliminaire resultaten van deze thesis zijn veelbelovend. Er werd een *in vitro* cell-gebaseerde assay ontwikkeld om het effect van PCDH10- en PCDH11X-expressie te evalueren. Deze assay kan gebruikt worden voor navolgende assays om de relevantie te testen van de moleculaire partnerinteracties die in dit doctoraat geïdentificeerd werden. Bovendien werden met behulp van de MAPPIT array interessante proteïnen geïdentificeerd die als vermeende interactiepartners van PCDHs fungeren. Tenslotte is in deze doctoraatsthesis de relevantie van deze interacties verder opgehelderd om zo de mechanismen gerelateerd tot PCDHs in gezonde en pathologische omstandigheden te kunnen identificeren.

INTRODUCTION

Chapter I. δ–PROTOCADHERINS

I.1 Protocadherins

I.1.1 Overview

The cadherin superfamily is a most interesting class of cell-cell adhesion molecules and the members are involved in creating and maintaining proper tissue architecture in embryonic development and tissue homeostasis[1–3]. Five major subfamilies form the cadherins superfamily, i.e. classical cadherins of type I, cadherins of type II, desmosomal cadherins, protocadherins, and cadherin-related molecules[4,5].

The classic cadherins show a single transmembrane domain and a cytoplasmic domain to which specific proteins bind collectively called catenins: p120-catenin (p120ctn) binds the cytoplasmic domain near the plasma membrane and β-catenin near the C-terminus. The cadherin complex is linked to the actin cytoskeleton via the α-catenin bind to β-catenin (**Figure 1**)[6]. A mechanotension function due to the interaction of cadherin-linked β-catenin/α-catenin complexes with the acto-myosin cytoskeleton is a unique characteristic of classic cadherins, if compared with other cell-cell adhesion molecules or even other members of the cadherin superfamily. This interaction is important for the role played by classic cadherins at cellular junctions, where the generation of force contributes critically to morphogenetic and physiologic processes[7,8].

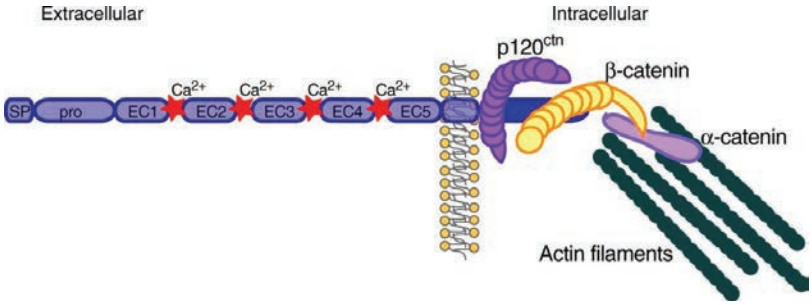


Figure 1 Cadherin domain structure.
Classic cadherins in vertebrates show 5 extracellular cadherin (EC1-5) repeat domains with EC domain interfaces mediated by calcium binding sites. The common interaction with catenins in their cytoplasmic domain is shown: this regulates cadherin functions and mediates interaction with the cytoskeleton and signaling molecules. All classical cadherins are synthesized with a signal peptide (SP) and pro-region (pro), which are removed during protein processing (modified from [9]).

Extracellular cadherin domains consist of calcium-binding repeats, referred to as ECs. The ectodomains of classical cadherins are responsible of their adhesive function as they mediate (generally homotypic) trans interactions between molecules on neighboring cells. Cis interactions can also occur via parallel interactions with like molecules on the same cell surface[10].

Cadherin-catenin complexes are dysregulated in various pathologies, notably cancer. In fact, the loss of function or decreased expression of any element of the cadherin-catenin-complex (CCC) leads to loss of the normal tissue architecture. Inappropriate regulation of their expression levels or functionality has been observed in human malignancies, in many cases leading to aggravated cancer cell invasion and metastasis[11,12].

The prototype cadherin is the Epithelial (E-)cadherin, a classical tumor suppressor that is generally downregulated in epithelial cancers leading to tumor growth and invasion[13]. It has been found to be mutated in lobular breast carcinoma and diffuse gastric carcinoma[12]. Different mechanisms depending on stage and type of tumor, can be responsible for the loss of E-cadherin expression, including genetic mutation but also hypermetylation of the promoter, silencing of the transcription, defective protein processing[12]. Another member of classic cadherin group, Neural cadherin (N-cadherin), is generally expressed by mesenchymal cells including neural cells. It has been shown repeatedly that cancer cells derived from epithelia inappropriately express N-cadherin. During the so-called epithelial-to-mesenchymal transition (EMT) a “cadherin switching” process occurs, where the loss of E-cadherin is related to an induction of the expression of N-cadherin with functional significance in cancer progression. The disease-related expression of N-cadherin contributes to cell motility and invasion of the malignant cells[9,12,14]. Many more cadherins have meanwhile been identified and an increasing number of them have been implicated in cancer as putative tumor suppressor or as a proto-oncogenic protein, and in various neurological disorders[6,11,15].

The protocadherin (PCDH) family represents the largest subgroup of the cadherin superfamily[5]. In humans, out of ∼110 cadherin superfamily members, 64 are protocadherins[15]. They are mainly but not exclusively expressed in the nervous system. PCDHs differ in various aspects from classic cadherins. The main differences with the cadherin family are the number of extracellular cadherin (EC) repeats (5 for cadherins and 6/7 for PCDHs) and the lack of conserved tryptophan residues, which are present at position 2 (W2) of the EC1 domain in classical type I cadherins, or at positions 2 and 4 (W2 and W4) of the EC1 of the type II cadherins. Those residues are responsible in classical cadherins for the formation of a hydrophobic pocket involved in homophilic binding. Since the residues and consequently the pockets are not conserved in PCDHs this might play a role in their

inability of forming strong homophilic bonds[16]. PCDHs can be grouped in clustered PCDHs (subdivided in PCDH-α, -β and -γ) and nonclustered PCDHs or δ-protocadherins[4,5]. The subfamily of nonclustered PCDHs is well conserved among vertebrates and comprises 10 members in humans. While evidence for cancer involvement of clustered PCDHs is sparse, accumulating evidence suggests that nonclustered PCDHs can function as tumor suppressors or proto-oncogenes in various non-neuronal human tumors[12]. (Reviewed in van Roy, 2014)[11].

I.1.2 History and Classification

When the study of the cadherin superfamily was still at his infancy, attention was also given to the identification of cadherin-related molecules such as desmoglein, desmocollins and T-cadherin. In this context, for example, Mahoney and colleagues showed in 1991[17] that the so-called *fat* tumor suppressor gene of *Drosophila* encodes a cadherin-related protein, a novel big transmembrane protein member of the cadherin type. Different studies have been performed to investigate the characteristics of the family and it has been shown that the cadherins and cadherin-related proteins are expressed in a large number of tissues in different organisms[17–19].

In 1993, Sano *et al.*[18] reported the first PCDH. They introduced the word *proto-* (from Greek πρῶτος, "first") to indicate them as "first cadherins". The study was carried out with the aim of identifying new members of the cadherin (super) family. To this end, degenerate primers for the extracellular repeats of classical cadherins were used in PCR experiments. The newly identified cadherin ectodomains were different from the ones of classical cadherins and, since then, isolated in a wide range of vertebrates and invertebrates pointing at an ancient, primordial cadherin motif. The family was then defined by distinct cadherin repeats in the extracellular domain, which were related to the primordial cadherin motif[18] (**Figure 2A**).

This study aimed at investigating which other molecules could cooperate with classic cadherins in cell-cell contact mechanisms. Pc42 and pc43, today known as PCDH1 and PCDH γC3, were identified first, but since then the number of the Pcdh family members increased dramatically: today more than 70 different *PCDH* genes are known in mammals. From this first study some peculiar characteristics of the members of the PCDH family were already disclosed. Quite typical is an overall structure similar to the one of the classic cadherins but with a completely different cytoplasmic domain. Moreover, despite the demonstration of Ca^{2+}-dependent cell adhesion activity, like in the case of classic cadherins, it became clear that the adhesion activity of PCDHs is weak as compared with the other members of the superfamily.

PCDHs resemble cadherins structurally: their ectodomains are composed of six or seven EC repeats, whereas classic cadherins have five. However, their cytoplasmic domains show no significant homology with those of classical cadherins. The family can be subdivided in two large groups based on the genomic structure and on sequence homologies: clustered (C) PCDHs (encoded by *PCDH-α, -β* and *-γ* gene clusters) and non-clustered (NC) PCDHs (δPCDHs and PCDH12 (**Figure 2B**))[4,5].

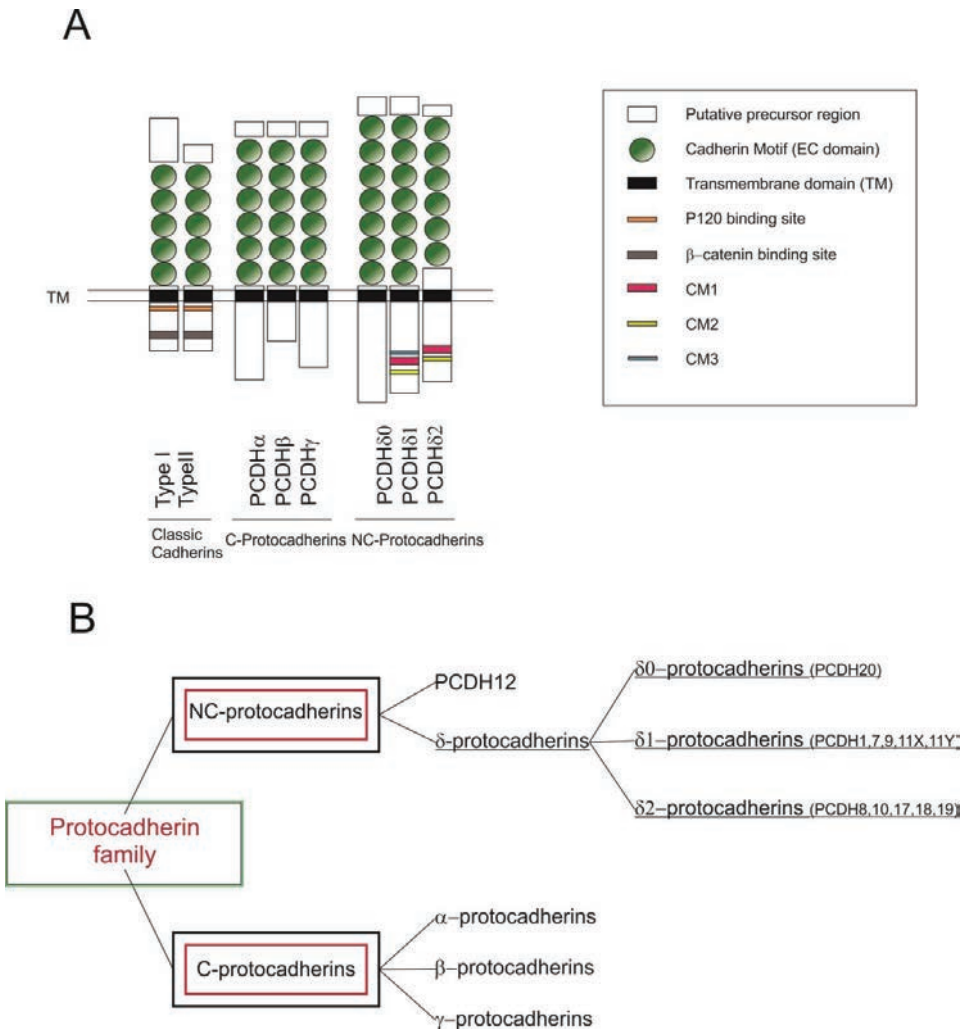


Figure 2 Structure and classification of the protocadherin family.

(A) Schematic overview of structures of the non-clustered and clustered protocadherins compared to the classic cadherins. Clustered cadherins show 5 EC domains in the extracellular part and p120-catenin and β-catenin binding sites in the intracellular domain. Protocadherins show 6 or 7 EC domains and the δ1 and δ2 subfamilies have 2 or 3 conserved motifs in the intracellular domain.

(B) Simplified phylogenetic tree of protocadherins from Hulpiau and van Roy, 2009 [5] *. Figure adapted from refs.*[4,5].

I.1.3 Molecular Evolution

The superfamily of Pcdhs has been studied in different ways to disclose their evolutionary origin. Considering an *Animalia* phylogenetic tree, the appearance of a Pcdh has been shown already in *Nematostella vectensis* (NvPcdh) (**Figure 3**). Analyzing the complete tree leads us to conclude that despite the existence of a single NvPcdh, the family mostly evolved in the Deuterostomia branch of the evolutionary tree. However, evidence is mounting that among the Protostomians also the Lophotrochozoans (e.g. mollusks, annelids) express one to numerous protocadherins, whereas the Ecdysozoans (e.g. arthropods, nematodes) have apparently lost the Pcdh genes ([5], and personal communication by Paco Hulpiau). In vertebrates, we see a maximum expansion of Pcdh genes, mainly due to generation of clustered *Pcdh* genes; in fact mammals, fishes and reptiles have between 60 and 80 Pcdh family members[20] .

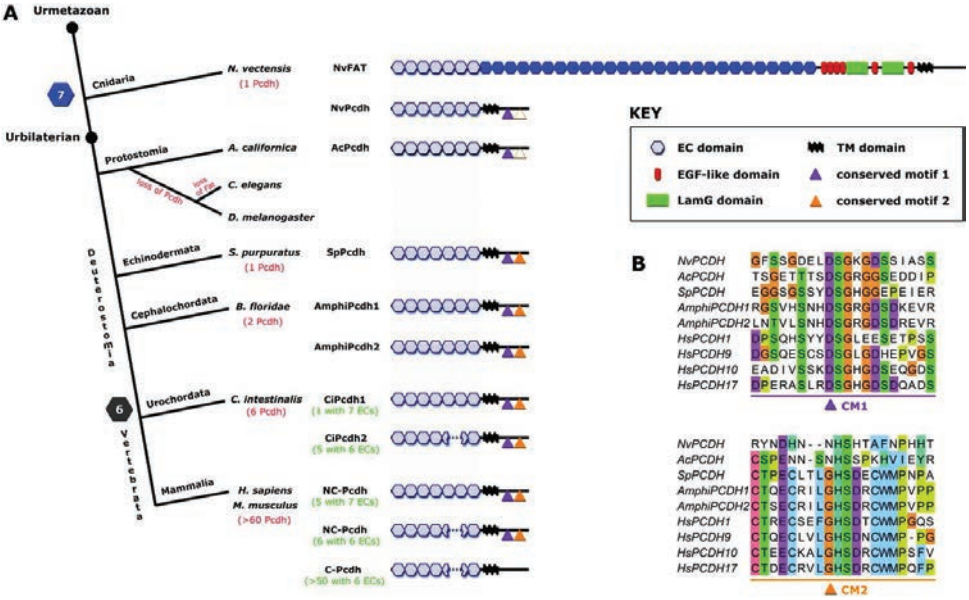


Figure 3 Metazoan evolution of protocadherins from the sea anemone to man.
(A) The cladogram shows relations among organisms for which protocadherin family members have been described[5,21]. On the right representative examples are shown. NvFAT is the FAT-related ancestor. The blue dotted lines show the internal deletion which reduce the number of EC domains from 7 to 6 in several modern Pcdhs. (B) Amino acid alignment of the conserved binding motifs 1 (CM1) and 2 (CM2) in the cytoplasmic domains of representative protocadherins. See text for more details (figure modified from[21]).

To investigate the evolutionary origin of protocadherins, different protein domains from several evolutionarily relevant species were compared. For this purpose, Hulpiau & van Roy [21] compared four different ancestral Pcdhs: 2 from the lancelet *Branchiostoma floridae* (AmphiPcdh1 and AmphiPcdh2), one from the sea anemone *Nematostella vectensis* (NvPCDH) and one from the California sea hare, *Aplysia californica*. This study revealed that all 4 have seven EC repeats and also conserved cytoplasmic motifs (CMs), highly homologous to those identified in the long isoforms of vertebrate δPcdhs[22]. From the analysis of the ectodomain of these ancient Pcdhs, one could propose that this ectodomain evolved from the N-terminal ECs of NvFAT protein. Previous studies based on the EC1 domains confirmed the close evolutionary relationship between FAT and protocadherins[5]. The loss of part of EC5 and EC6 from a 7EC-Pcdh most likely is responsible for the arising of the 6EC-Pcdh subfamily. Finally, it is conceivable that the numerous vertebrate C-Pcdhs with 6 EC domains were originating from ancestral NC-Pcdhs with 6EC-domains[21].

The molecular evolution of the cadherin superfamily (and consequently of the Pcdh family) has been analyzed by comparing the EC1 domains only (while confirmed in a EC1-to-EC7 analysis)[5]. A circular tree of the cadherin superfamily based on these EC1 sequence homologies was generated and showed a very clear dissection into two major branches in an evolutionary context [5]. The so-called EC1 based cadherin major branch could again be subdivided into two groups: C-1 (which includes type I and type II cadherins, the desmogleins, the desmocollins and the 7D-cadherin family) and C-2 (composed by type III cadherins, the FLAMINGO/CELSR cadherins and type IV cadherins). The second major branch is called EC1-based cadherin-related major branch (EC1-CrMB) and is composed of ten families including Protocadherins, which constitute the largest cadherin family, and has been annotated as the Cr-1a sub branch. C-Pcdh and NC-Pcdhs are taking obviously distinct positions in the EC1 based phylogenetic tree: Pcdh1, Pcdh7, Pcdh9, Pcdh11 (all delta1 Pcdhs) and Pcdh20 (δ0) form a compact subbranch, while δ2Pcdhs are spread over different subbranches. Having 6 EC domains, they appear to be closer to C-Pcdhs than to δ1Pcdhs. Human PCDH10 and PCDH11 show an intriguing evolutionary position within the family members. In the analysis based on EC1 domains, PCDH10 surprisingly localizes within the C-Pcdh subbranch, suggesting itself as a bridge between clustered and non-clustered Pcdhs. Concerning PCDH11, the X- and Y-linked genes which encode PCDH11X and PCDH11Y proteins, respectively, share up to 98.3% amino acid sequence identity, which means that they have evolutionarily diverged quite recently.

Another analysis comparing cytoplasmic domains of protocadherins was performed showing, as expected, great homology between the three groups of C-Pcdh, and also within the NC-Pcdh

subgroup, but no homology at all between the cytoplasmic domain of, respectively, C- and NC Pcdhs[5]. Results are shown in the following circular phylogenetic trees (**Figure 4**):

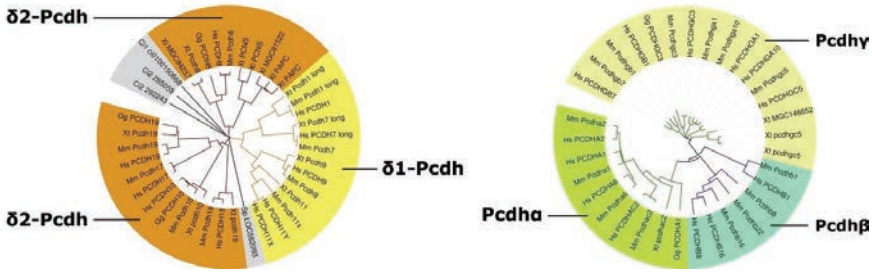


Figure 4 Circular phylogenetic trees of selected members of the protocadherin family.
The analysis is based on homologies in the cytoplasmic domain. Left: C-Pcdhs, Right NC-Pcdhs (modified from[5]).

I.2 Clustered Protocadherins

With 50 members, clustered Pcdhs are the largest Pcdh group, and were first classified as cadherin neuronal receptor (CNR) gene family in mouse brain. The many clustered *Pcdh* genes have been found and analyzed in several vertebrate species, including rat, mouse, chicken, zebrafish and man. In mammals, the C-PCDH family is divided into three gene clusters: PCDH-α, -β and –γ. The main characteristic of the C-PCDH gene clusters is a repetition of single large exons encoding multiple variants of the (ectodomain + TM + part of the cytoplasmic domain), in combination with shared smaller exons, encoding the remaining apart of the cytoplasmic domain (**Figure 5**).

I.2.1 Genomic organization and protein structure of clustered protocadherin

Concerning the genomic organization of *C-Pcdh* genes, the main characteristic is the presence of a region formed of a tandem array of variable exons and a constant region formed by a set of shared exons. This constant region (missing in the Pcdh-β subgroup) has been shown to be highly conserved among vertebrates. The variable exons are homologous among the Pcdh-α, -β and –γ groups and encode each time for the six N-terminal EC-like domains, the transmembrane part and a short component of the cytoplasmic domain. The structural characterization of the protocadherin EC1 domain revealed crucial variations including two specific motifs. The first one is a RGD (Arg-Gly-Asp)

motif well conserved in Pcdh-α proteins, PCDH17, PCDH19 and CDHR5. Such motif has been shown to be involved in integrin-dependent cell adhesion: it is recognized by integrins on their ligands, such as fibronectin, fibrinogen, von Willebrand factor and many other large glycoproteins[23]. The other very well conserved sequence in the EC1 domain is a Cys-$(X)_5$-Cys motif, present in all clustered PCDHs and in δ2PCDHs but not in classical cadherins. Due to the ability of forming disulfide bounds and the probable contribution to formation of cis-tetramers between PCDHs, it could be considered as a novel adhesion interface.

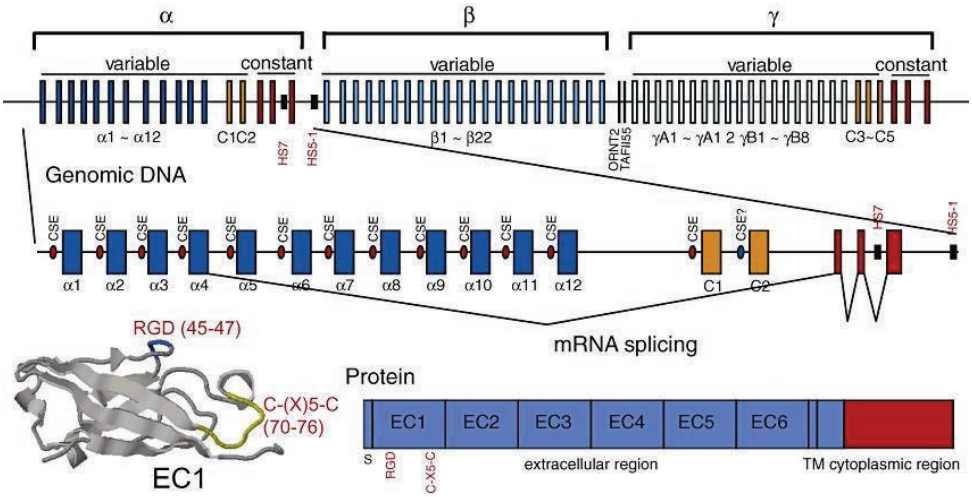


Figure 5 Genomic organization of the mouse clustered Pcdh genes.
The three clusters of mouse Pcdh genes, Pcdh-α (α), Pcdh-β (β), and Pcdh-γ (γ), are arranged in chromosome 18 of the mouse. The Pcdh-α and Pcdh-γ gene clusters include variable and constant regions, and C-type variable exons: Pcdh-αC1 and -αC2, and Pcdh-γC3, -γC4, and -γC5. The conserved sequence element (CSE) is located upstream of each transcription start site and is part of the promoter region except for the Pcdh-αC2 promoter. The combination of one of the variable regions with three constant region exons produces by cis-splicing a mature mRNA. Pcdh-α and -γ constant exons encode large parts of the cytoplasmic tails. The small black boxes indicate identified DNase I hypersensitive enhancer sites (HS7 and HS5-1) in the Pcdh-α cluster. The ribbon diagram of the represented NMR structure of the EC1 domain of Pcdh-α4 is modified from[24]. Picture modified after[25].

I.2.2 Gene expression and protein interactions

The gene expression of clustered PCDHs shows a unique mechanism due to the presence of multiple alternative promoters, responsible for high molecular diversity in individual neurons. Their expression is monoallelic and shows a combinatorial regulation, with the exception of the C-type

isoforms[a]. This was shown in a single-cell analysis of Purkinje cells using multiple RT-PCR reactions[26]. There are different kinds of monoallelic gene expression patterns: random X chromosome inactivation in females, imprinting genes from the maternal or the paternal allele or random monoallelic gene expression in autosomes. The regulation of clustered-PCDH expression was found to follow this last mechanism[16]. The genomic organization of *PCDH-α* and *PCDH-β* gene clusters resembles the organization of the immunoglobulin and T-cell receptor gene clusters, which are characterized by generation of a great variability in antibody/TCR production. For clustered PCDHs, related but novel mechanisms yield similar results. Different models were proposed to explain the genomic regulation of clustered *PCDH* genes and in 2002 two different groups independently showed that their transcription starts from multiple promoters upstream of each variable exon and that processing of the mRNA occurs via cis-splicing (see also **Figure 6**). Trans-splicing is also possible but uncommon. This shows that the expression regulation already starts at the transcriptional level and that differential promoter activation is the key for the diversity of the expression in neurons. The result is a sophisticated mechanism for generating single cell diversity in the brain and for the production of an important number of isoforms with different extracellular domain sequences[27–29]. Follow up studies, which included the entire protocadherin γ cluster (Pcdh γ A and Pcdh γ B) and the C-type variable exons (C1-C5) of the protocadherin α cluster, showed something even more surprising. All the C-type isoforms showed biallelic expression. These results indicate that the Pcdh family shows differential allelic gene regulation in the same gene cluster, which is a unique characteristic of the family[30].

[a] As it is investigated by phylogenetic studies, PCDH isoforms show high homology between each other; despite this, there are exceptions represented by specific components of the PCDH-α gene cluster (PCDHαc1 and PCDHαc2) and the PCDH-γ gene cluster (PCDHγc3, PCDHγc4 and PCDHγc5). Those 5 PCDHs are the so-called C-type isoforms, which evolutionary diverge from the other members of the family but are similar to each other.

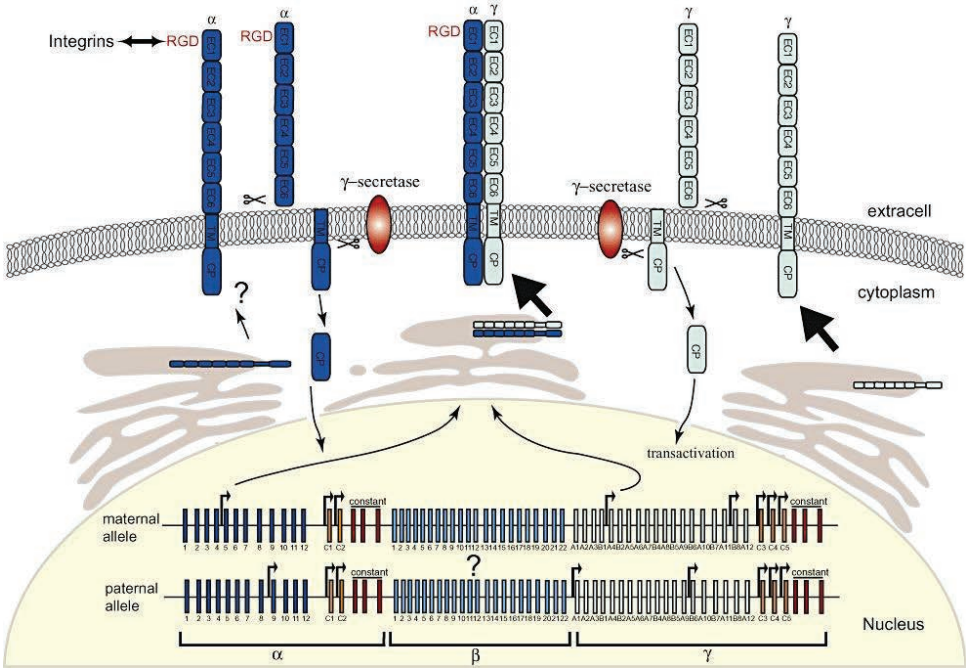


Figure 6 Schematic overview of the gene regulation and protein interaction of Pcdh-α and Pcdh-γ.
In the nucleus, monoallelic and biallelic types of gene regulation in the Pcdh-α and Pcdh-γ clusters can occur in a single neuron. In each maternal or paternal allele, the Pcdh-α1 to α12, Pcdh-γA1 to γA12, and Pcdh-γB1 to γB8 variable exons exhibit monoallelic and combinatorial expression. By contrast, the C-types of the Pcdh−α and Pcdh-γ variable exons are expressed biallelically. This model is based on expression data obtained from single Purkinje cells. Arrows indicate the strongly expressed variable exons. The differential expression of these genes in individual neurons may specify neuronal identity in the brain. Outside the nucleus, the co-expression of the Pcdh-α (dark blue) and Pcdh−γ (light green) proteins greatly enhances their transportation from the endoplasmic reticulum to the cell surface, compared with the expression of Pcdh-α alone. A portion of the Pcdh-α protein forms a heterocomplex with Pcdh-γ. These protein complexes increase the combinatorial diversity on the cell surface, and contribute to the differential characteristics specifying the individuality of single neurons. The Pcdh-α protein interacts with integrins through its RGD motif. The Pcdh-α and Pcdh-γ cytoplasmic domains are cleaved by γ-secretase after the extracellular domain is cleaved by metalloproteinase. Their C-terminal fragments are transported into the nucleus, and then the Pcdh-γ fragment may transactivate the Pcdh-γ genes. The formation of cis-heteromers between Pcdh-α and Pcdh-γ alters susceptibility to this processing by proteolysis. α, Pcdh-α; γ, Pcdh-γ; EC, extracellular cadherin domain; TM, transmembrane domain; CP, cytoplasmic domain; ER, endoplasmic reticulum. (Figure from[25]).

The expression of C-PCDHs is responsible for the regulation of neuronal development in vertebrates. Different mechanisms concur to regulate neural circuit formation through specific cell-cell interactions. A self-avoidance mechanism is responsible for the repulsion between sibling dendritic branches that arise from a single soma: this process is not shown for every kind of neurons but in mammals it has been documented in Purkinje cells and some types of retinal cells[31–33] (**Figure 7**). By the self-avoidance mechanism, mutual interaction of dendrites of the same neuron is minimized, while the same dendrites can interact with the ones of other neurons of the same kind. This process is called self/non-self discrimination.

C-PCDHs have been shown to play a role in both self-avoidance and self/non-self discrimination[33–37], and functional studies have been performed for the PCDH-α and Pcdhγ gene clusters[38].

Lefebvre[33] and colleagues showed the roles of *C-Pcdh* genes in self avoidance and self/non-self discrimination processes using retinal starburst amacrine cells (SACs) and cerebellar Purkinje cells, which express Pcdh-γ genes and show self-avoidance. Self avoidance was lost if Pcdhs were removed and it resulted in the formation of self synapses (autapses) in SACs due to their ability of forming now dendro-dendritic synapses. If only one Pcdh isoform was re-expressed, self-avoidance was reconstituted but not self/non-self mechanism: this indicates that the diversity is the base of interaction between proximity neurons[33,37].

Besides cell autonomous self avoidance, there are other mechanisms responsible for the C-Pcdh-dependent regulation of neuronal development. Different studies carried out in other CNS cell lines have demonstrated non autonomous roles played by the Pcdh-γ members via heteroneuronal and neuron-astrocyte interactions. Using a manipulated set of Pcdh-γ isoforms in the cerebral cortex, for instance, the dendrite arborisation has been shown to be promoted by

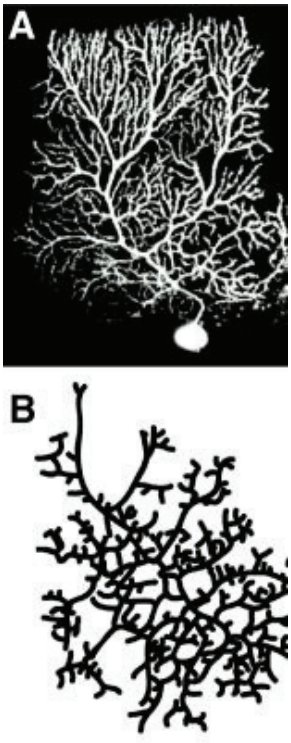


Figure 7 Isoneuronal neurites repel one another. (*A) The branches of a mouse cerebellar Purkinje cell dendritic arbor, almost never overlap with one another (B) The branches of a retinal ganglion cell in culture almost never overlap with one another. Modified from*[32].

the increased probability of homophilic matching between a neuron and neighboring neurons. If the trans-interaction is instead compromised by mismatching of the isoforms, the arborisation cannot develop and the result is similar to the one seen in Pcdh-γ KO neurons where self avoidance is impaired[33].

An *in vitro* study on leukemia cells K562, which do not express endogenous PCDH-γ, has revealed the existence of strictly homophilic trans interactions for PCDH-γ[39]. Compared to classical cadherins, PCDHs show more weak trans-homophilic adhesion activity. In a protein coated bead aggregation assay, a mutated neural cadherin (NCAD), where the EC1 domain (responsible for the interaction) was replaced with the EC1 domain of CNR/PCDH-α4, loses its adhesive property, meaning that the difference in structure of the EC1 domain is the main responsible for the weak trans-homophilic adhesion activity of PCDHs[24]. In contrast, cis homodimers[40,41] and cis heterodimers[42] of C-PCDHs have been observed on the cell surface and it has been proposed that PCDHs acquire the trans-homophilic adhesion activity after the formation of complexes with other PCDHs. Cis-tetramers have been shown to be the unit for homophilic trans interaction by C-PCDHs: the interactions involve different isoforms producing promiscuous cis-multimers with no isoform specificity. For the 22 possible isoforms of PCDH-γs, about 234,256 distinct interacting surfaces can be predicted, which offers an enormous diversity of possible interactions on the cell surface [39]. PCDH-γ have been shown to form complexes also with PCDH-α and PCDH-β isoforms[43].

Cells expressing different combinations of the 3 groups of C-PCDHs have been shown to aggregate through homophilic trans-interactions between the specific PCDHs expressed, but they fail to aggregate if the combination differs even for one isoform in the combination, showing that cell-cell adhesion will occur only between cells expressing exactly the same combination of C-PCDHs (**Figure 8**)[44].

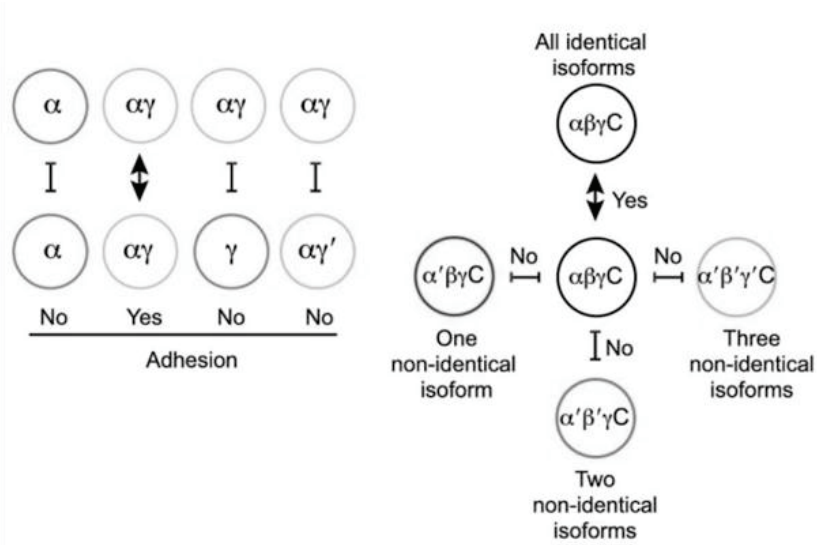


Figure 8 Cell–cell adhesion is induced only between cells expressing the same combination of clustered protocadherins

The figure on the left shows that only the same combination of isoforms can lead to cell-cell adhesion. Pcdh-α isoform alone does not induce adhesion, probably because in this context it is not recruited at the membrane. On the right, it is shown that Pcdh-α, Pcdh-β, Pcdh-γ and the C isoforms interact with a cell with the same combination of isoforms: one mismatch is enough to compromise the adhesion. (Modified after [45]).

For C-PCDHs, the EC1 domain is not the responsible region for the specificity of the interaction while the specificity can be mediated by EC2/3[44,46,47]. Two independent studies have reported the crystal structure of mouse α-, β-, and γ-Pcdh fragments which include EC1–3[47,48]: those fragments were monomeric in solution and both studies indicated that in contrast to classical cadherins, the trans adhesion between cells depends on the EC1–EC4 domains. Cell aggregation experiments in human immortalized myelogenous leukemia line (K562) have shown that C-PCDHs interact in antiparallel orientation through the formation of zipper-like cis homo- or hetero-dimeric assemblies.

In addition to their cell-cell adhesion activity, clustered PCDHs play a role in mediating intracellular signaling. Their intracellular domain is different from the one of the classical cadherins, and lacks the conserved catenin binding sites. The cytoplasmic domains of PCDH-α and PCDH-γ interact with two tyrosine kinases, proline-rich tyrosine kinase 2 PYK2 and focal adhesion kinase (FAK)[49]. These interactions result in the inhibition of the kinase activities *in vitro* and *in vivo*. The activity of PYK2 is upregulated in neurons of Pcdh-γ KO mice. Studies conducted in mice, zebrafish and chicken have

shown that overexpression of PYK2 or ablation of PCDH-γ induce apoptosis of neurons[49,50] indicating that the interaction correlates with the cell survival of a subset of neurons. The interaction with FAK is also responsible for positive regulation of dendritic arborisation in cortical neurons[46]: in neurons lacking Pcdh-γ the dendrite arborisation is disrupted due to over activity of FAK and PKC. This function is regulated by PKC mediated phosphorylation of a C-terminal serine (Ser-922) within the Pcdh- γ constant domain[51].

Furthermore, both PCDH-α and PCDH-γ proteins bind the receptor tyrosine kinase Ret, which is required for the stabilization and phosphorylation of their intracellular domains. Several more proteins have been reported to interact with protocadherins, and include phosphatases, kinases, adhesion molecules and synaptic proteins. Taken together these findings suggest that PCDHs can form big heteromeric complexes with very different functions, which need to be studied in more detail in order to be fully elucidated.

I.3 δ-Protocadherins

In man and mouse, the δ-Protocadherin subfamily comprises 10 members, which can be subdivided into δ0PCDH (one member, PCDH20), δ1PCDHs (four members: PCDH1, PCDH7, PCDH9 and PCDH11X/Y), and δ2PCDHs (five members: PCDH8, PCDH10, PCDH17, PCDH18 and PCDH19) (**Table 1**). This subdivision is based on overall sequence homology, number of EC repeats (seven versus six), and conservation of specific amino acid motifs in cytoplasmic domains. The δ1 and δ2PCDHs are expressed as both short and long isoforms, which are generated by alternative splicing and differ from each other in the size of their cytoplasmic domain. All the long isoforms show highly conserved motifs in their cytoplasmic domains (CM1 and CM2 for δ1 and δ2PCDHs; plus an additional CM3 for δ1PCDHs only) (**Figure 2**)[4].

Table 1 Human δPcdhs family members

HGNC ID	Approved Symbol	Approved Name	Previous Symbols	Synonyms	Chromosome
8655	PCDH1	protocadherin 1		pc42	5q31.3
13404	PCDH10	protocadherin 10		OL-PCDH, KIAA1400	4q28.3
8656	PCDH11X	protocadherin 11 X-linked	PCDH11	PCDH-X, PCDHX, PPP1R119	Xq21.3
15813	PCDH11Y	protocadherin 11 Y-linked	PCDH22	PCDHY	Yp11.2
8657	PCDH12	protocadherin 12		VE-cadherin-2	5q31.3
14267	PCDH17	protocadherin 17		PCDH68, PCH68	13q21.1
14268	PCDH18	protocadherin 18		KIAA1562, PCDH68L	4q28.3
14270	PCDH19	protocadherin 19	EFMR	KIAA1313, EIEE9	Xq22.1
14257	PCDH20	protocadherin 20		PCDH13, FLJ22218	13q21
8659	PCDH7	protocadherin 7		BH-Pcdh, PPP1R120	4p15
8660	PCDH8	protocadherin 8		PAPC, ARCADLIN	13q21.1
8661	PCDH9	protocadherin 9			13q21.32

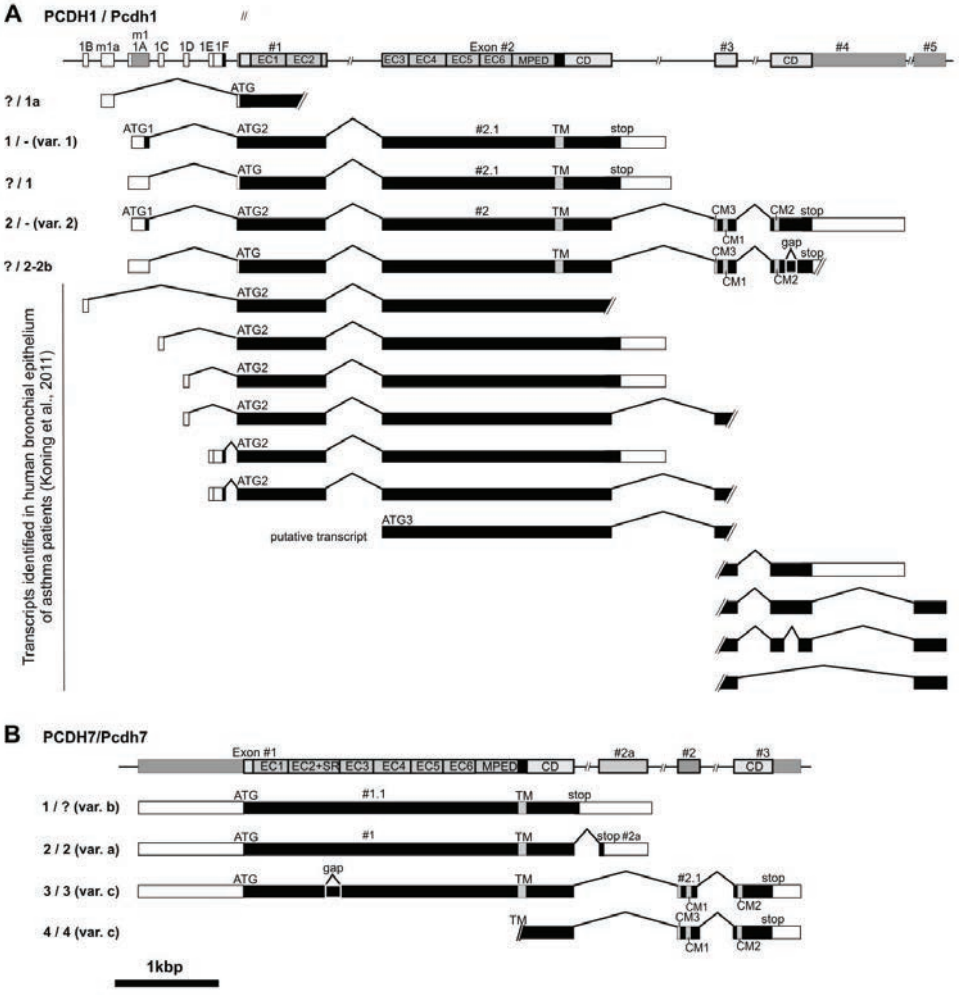
I.3.1 Genomic organization and protein structure of δPCDH

The genomic organization of non-clustered protocadherin genes is peculiar and shows obvious differences with the classic cadherin gene structure. The first main difference between the two subfamilies of cadherins is the way in which the EC domains are encoded: in protocadherins (both clustered and non-clustered) a single large exon is responsible for encoding the extracellular domain,

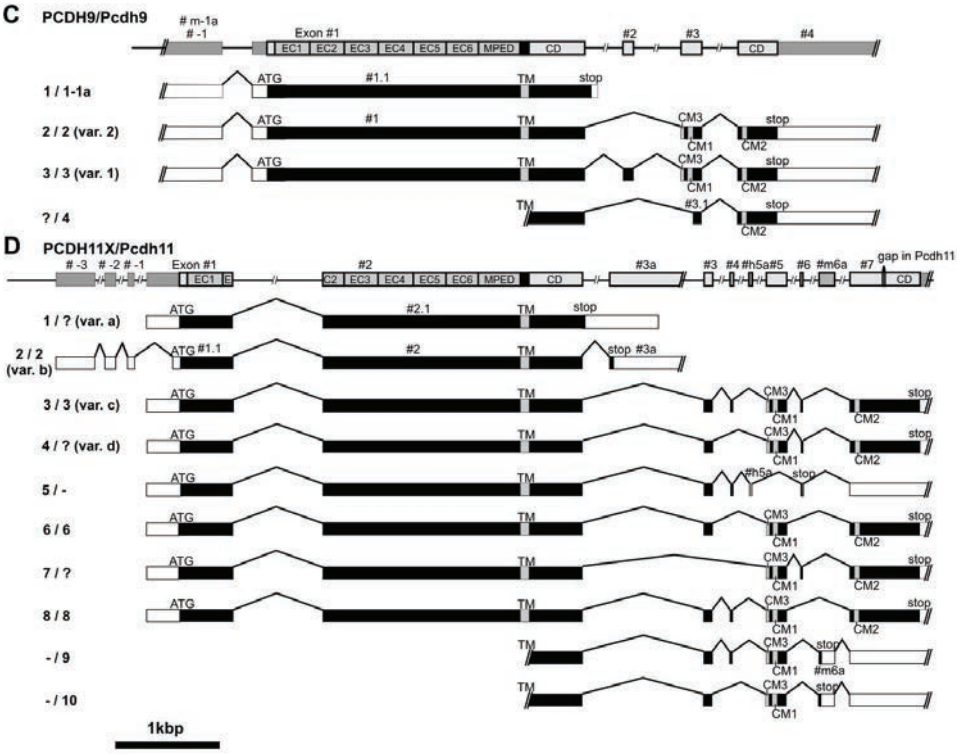
the transmembrane domain and part of the cytoplasmic one, while for classical cadherins there are many smaller exons involved. Another difference is the lack of a prodomain[b] in Pcdhs, while this is encoded by classical cadherin genes. *δPcdhs* genes are organized in the same way as the other members of the *Pcdh* gene family but show also additional unique characteristics concerning their cytoplasmic domain; this differs not only from classical cadherins but also from C-PCDHs due to the presence of conserved motifs (CM1-CM3). CM1 and, when it is present, also CM3 are encoded by the same exon, while CM2 is encoded by an exon encoding also the 3'-UTR. Alternative splicing occurs often and also intraexonic splicing occurs, which produces a large number of protein isoforms (**Figure 9**). Interestingly, different PCDHs show different degrees of splice variation. PCDH11 X/Y (**Figure 9D**) are expressed in a large number of different isoforms and this is due to the variety of transcripts, using various short exons from a larger set included in their gene. The 'long' isoform of PCDH11X shows a sequence of 11 aa repeated 9 times, and comprising two proline residues per repeat: this is the region where PCDH11X and PCDH11Y show the biggest difference[52]. PCDH11Y is encoded by the human genome only (chromosome Y), due to a recent duplication from the X-chromosome encoded PCDH11X which occurred after the divergence between chimpanzees and humans[53,54]. Only two isoforms are known in human for PCDH10 proteins, called 'long' and 'short' (**Figure 9F**). If no transcript splicing occurs the short isoform is produced, which shows a short cytoplasmic tail in comparison with the long isoform and which lacks the two CMs. The PCDH10 long isoform, generated by alternative mRNA splicing, includes both CMs in its elongated cytoplasmic domain. In mice, *Pcdh10* produces four protein isoforms. PCDH1 (**Figure 9A**) shows a lot of different transcripts in primary human bronchial epithelial cells including a new isoform lacking almost the entire extracellular domain: such a large number of different transcripts in the same cell type may be an indication of a big variety of functions covered by the different isoforms[55]. Furthermore, all δPcdhs share the presence of a cysteine at the beginning of the cytoplasmic domain[5]. In PCDH1 and other PCDHs this cysteine is palmitoylated which might be necessary for protein trafficking and compartmentalization in polarized neurons[56]. If the modification does not occur, PCDH1 is not localized at the plasma membrane (Karl Vandepoele, not published), but when only this particular cysteine residue is mutated, the protein localization is still correct pointing at a redundancy of residues important for this specific function.

[b] Prodomain: classical cadherins are activated through cleavage in the late Golgi of a prosequence. This domain corresponds to the folded region of the prosequence, and is termed the prodomain. The prodomain shows structural resemblance to the cadherin domain, but lacks all the features known to be important for cadherin-cadherin interactions[215].

δ1-Pcdhs

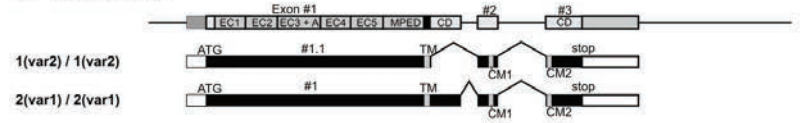

A
PCDH1 / Pcdh1
1B m1a m1 1A 1C 1D 1E1F
#1
EC1 EC2
Exon #2
EC3 EC4 EC5 EC6 MPED CD
#3
CD
#4
#5
? / 1a
ATG
1 / - (var. 1)
ATG1
ATG2
#2.1
TM
stop
? / 1
2 / - (var. 2)
#2
CM3
CM2 stop
CM1
? / 2-2b
gap
Transcripts identified in human bronchial epithelium of asthma patients (Koning et al., 2011)
putative transcript
ATG3
B
PCDH7/Pcdh7
Exon #1
EC1 EC2•SR EC3 EC4 EC5 EC6 MPED CD
#2a
#2
#3
1 / ? (var. b)
#1.1
2 / 2 (var. a)
#1
stop #2a
3 / 3 (var. c)
4 / 4 (var. c)
1kbp

δ1-Pcdhs (cont'd)

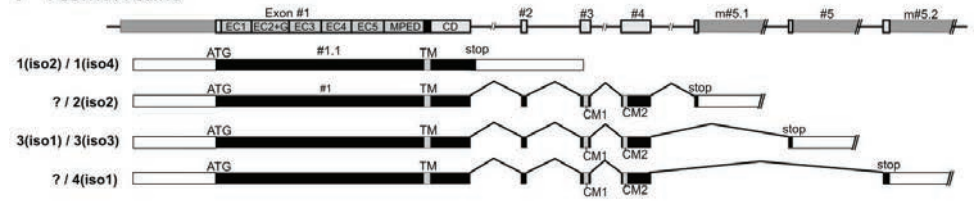


δ2-Pcdhs

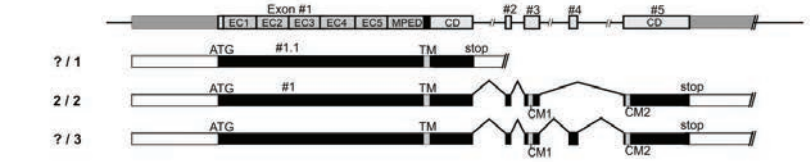
E PCDH8 / Pcdh8



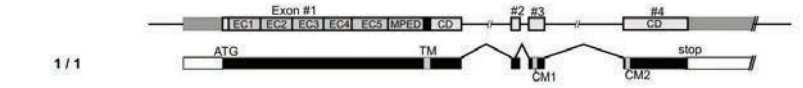
F PCDH10 / Pcdh10



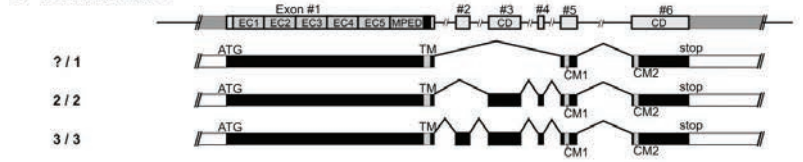
G PCDH17 / Pcdh17



H PCDH18 / Pcdh18

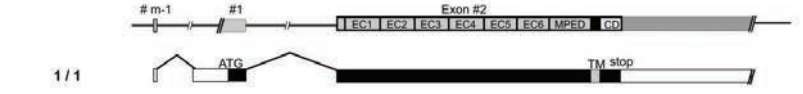


I PCDH19/Pcdh19



δ0-Pcdh

J PCDH20 / Pcdh20



1kbp

Figure 9 Gene structure and alternative transcripts of the δ-protocadherin family in man and mouse.
Results are basically identical for human (isoform numbers before slash signs) and mouse (isoform numbers after slash signs), although some significant differences exist. If a particular transcript does not exist (or has not been identified yet) in one of the two species, this is indicated by a minus sign before or after the slash sign. Variant annotations between parentheses refer to the literature, but lack overall consistency. Question marks denote presumable transcripts that need to be validated experimentally in one of the two species. All transcripts shown are protein coding. For each transcript the protein coding part is indicated by black bars, whereas UTRs are indicated by open bars, and splicing events by an angled line. Exons are drawn to scale and numbered as indicated. Exon #1 is defined as the exon that contains the standard start codon (indicated by ATG); alternatively used exons are indicated by the extension 'a'; exons that are elongated by read-through in the neighboring intron or that become shortened by an intraexonic splice acceptor are indicated by the extension '.1'. Exons that occur only in either the human or the mouse genome are indicated by 'h' and 'm', respectively. Gray boxes represent exons; the intervening line represents intronic sequence (the latter not at scale). Black boxes represent coding sequences in the transcripts, white boxes represent 5' or 3' untranslated regions. A, alanine-rich region; CM1, CM2 and CM3, conserved motifs; CD, cytoplasmic domain; EC, sequences encoding the extracellular cadherin repeats; MPED, Membrane proximal extracellular domain; G, glycine-rich region; SR, serine-rich region; TM, sequence encoding the transmembrane domain. (A-D) δ1-protocadherins. (A) Protocadherin-1 genes (human PCDH1, mouse Pcdh1) and derived transcripts: 1, truncated transcript; 2, long transcript. According to the literature[57], the human protein starts with the ATG in exon #1a, which is not conserved in mouse or other species. In contrast, the first in-frame ATG in exon #1 is conserved across many species including human and shows a better Kozak consensus sequence[22]. Recently, various novel PCDH1 transcripts have been identified in primary bronchial epithelium of patients with asthma[58]. Novel exons were identified on the 5' end of PCDH1, which were connected to the short or long isoform or both. Variations on the 3' end of PCDH1 can result in alternative usage of CM2. (B) Protocadherin-7 genes (human PCDH7, mouse Pcdh7) and derived transcripts: 1 and 2, two truncated transcripts, only differing at the 3' end; 3, differs from the transcript 4 by a shortened exon 2. (C) Protocadherin-9 genes (human PCDH9, mouse Pcdh9) and derived transcripts: 1, short transcript; 2, long transcript; 3, transcript with additional exon #2; 4, transcript with variant exon #3.1, lacking CM3 and CM1. (D) Protocadherin-11 genes (human PCDH11X, mouse Pcdh11) and derived transcripts: 1 or a, short CP domain; 2 or b, use of alternative exon #3a; 3 or c, full-length transcript; 4 or d, idem without exon #4; 5, use of alternative human exon #h5a resulting in frameshift; 6, neither exon #4 nor exon #6; 7, neither exon #3 nor exon #4; 8, no exon #6; 9, use of alternative mouse exon #m6a, resulting in premature stop codon and loss of CM2; 10, as transcript 9 but with exon #4 skipped. The human paralog PCDH11Y is not depicted here. (E-I) δ2-protocadherins. (E) Protocadherin-8 genes (human PCDH8, mouse Pcdh8) and derived transcripts: transcript 1 differs from transcript 2 by a shortened exon #1. (F) Protocadherin-10 genes (human PCDH10, mouse Pcdh10) and derived transcripts: 1, short transcript; 2-4, long transcripts only differing at their 3' end. (G) Protocadherin-17 genes (human PCDH17, mouse Pcdh17) and derived transcripts: 1, short transcript; 2, long transcript; 3, transcript with additional exon #4. (H) Protocadherin-18 genes (human PCDH18, mouse Pcdh18) and derived transcript: 1, long transcript. (I) Protocadherin-19 genes (human PCDH19, mouse Pcdh19) and derived transcripts: 1, transcript skipping exons #2, #3 and #4; 2, transcript with additional exons #3 and #4; 3, idem with additional exon #2. (J) δ0-protocadherin. Protocadherin-20 genes (human PCDH20, mouse Pcdh20) and derived transcript: 1, transcript with short CP, lacking CM1, CM2 and CM3. (Figure from Kahr I., PhD Thesis, adapted from[22]). Assembly releases of the human genome (GRC37/hg19) and of the mouse genome (GRCm38/mm10), available at UCSC Genome Browser Gateway (http://genome-euro.ucsc.edu/), further extended with unpublished data obtained in van Roy's laboratory.

I.3.2 Expression patterns

δ-Protocadherins, like other Pcdhs, are mainly expressed in the nervous system but most (if not all) of them are also expressed in non-neuronal tissues[4]. Considering time and space, each member of the family shows a specific expression pattern. As already mentioned, most δPcdhs are expressed as different isoforms but generally, published studies do not discriminate between them. The pattern of expression differs considerably in different models for each Pcdh. We might use the expression pattern of Pcdh10 as a representative example: PCDH10 is expressed mainly in brain and in the nervous system in general, including the olfactory and the limbic system (therefrom their original name OL-PCDH)[59], but different *PCDH10* mRNA expression levels can be found also in most of the fetal and adult tissues in humans[60,61]; in contrast in mouse, Pcdh10 is expressed mainly in brain, spinal cord and eye[62]. The spatiotemporal differences in the expression of the various family members have been shown in a representative study in chicken cochleae[63]: at embryonic day 3 (E3) the first set of δPcdhs can be identified. From E11 on, various Pcdhs show different localizations on the discrete cell types of the cochleae: Pcdh1 is expressed by spindle-shaped cells and acoustic ganglion cells; Pcdh7 and Pcdh17 are exclusively expressed in supporting cells, cuboidal cells, hyaline cells and acoustic ganglion cells, while Pcdh9 is expressed by acoustic ganglion cells; Pcdh8 is expressed in hair cells, spindle-shaped cells and acoustic ganglion cells; Pcdh10 mRNA is expressed in spindle-shaped cells and acoustic ganglion cells at later stages. mRNAs of Pcdh1, Pcdh18 and Pcdh19 are also expressed in blood vessels of the cochlea[64]. Taken together those data indicate that the differential expression of several δPcdhs in the embryonic and fetal stages in different species might play important functional roles in embryonic development and tissue morphogenesis. In the nervous system, the main expression system common for all members of the δPcdh family, different studies have shown a large range of different distribution patterns of δPcdhs in time and space in the brain. Pcdh expression changes following the phases of brain development and it is specific for the different layers in the cerebral cortex[65,66]. The adhesion properties and restricted expression patterns of δPcdhs have led to the speculation that they are involved in developmental processes such as brain nucleus formation, axon migration, and synapse formation[15]. The complexity of neuronal identity and neuronal networks in the cerebral cortex may be due to the combinatorial occurrence in single neurons of discrete cadherin and protocadherin molecules. Indeed, combinatorial expression of molecules belonging to the same cadherin family can contribute to the molecular specification of individual cells and their mutual connections. Krishna and colleagues demonstrated a combinatorial molecular code in single neurons by checking the cadherin expression in the somatosensory cortex of the adult mouse[67]. δPcdhs show the same combinatorial behavior

also in the subregions of adult hippocampus and dentate gyrus along the septotemporal axis: the expression patterns of Pcdh1, Pcdh9, Pcdh10 and Pcdh20 showed septal preferences, whereas the expression patterns of Pcdh8, Pcdh11, Pcdh17 and Pcdh19 showed temporal preferences, pointing at their involvement in the formation of hippocampal circuits[68]. Concerning previous stages, such as early post-natal stages, other studies showed region specific patterns of δPcdh expression in developing rat brain[66]. NC-Pcdhs were analyzed in rat brain using *in situ* hybridization. Pcdh7 and Pcdh20 mRNAs have a common expression pattern in the somatosensory and visual cortices, Pcdh11 and Pcdh17 mRNAs are mostly expressed in the motor and auditory cortices, and Pcdh9 mRNA shows higher expression in the motor and main somatosensory cortices, confirming their region specific expression patterns. This study was performed in the cerebral cortex during the early postnatal stage (P3), which is relevant and crucial for the network formation, and suggests a role for δPcdhs in the establishment of selective synaptic connections[66].

I.3.3 δPcdhs and cell-cell adhesion

δPcdhs show only a weak adhesion function: Ca^{2+}-dependent homophilic cell-cell adhesion is occurring but much weaker than the one achieved by classical cadherins. Pcdh1, Pcdh8, and Pcdh18a have been shown to affect cell adhesion using cell sorting assays while Pcdh7, Pcdh10, Pcdh19, and Pcdh17 has been shown to mediate adhesion in cell aggregation assays[69–74].

Few examples of homophilic adhesion have been reported for *in vitro* studies on δPcdhs, for instance for Pcdh1/AXPC[18,70], Pcdh7/NFPC[75], Pcdh8/Arcadlin/PAPC[71,76], Pcdh10[72] and Pcdh19[77]. The preference of homophilic interaction instead of a heterophilic one has been used to mediate cell-sorting in cell culture and *in vivo*. The cytoplasmic domain was found to be necessary for Pcdh7 functionality in the epidermis of Xenopus[75], but this was not the case for Pcdh19[77] and PAPC[78], which form oligomers via the conserved cysteine residues in their ectodomain. Pcdh1/AXPC and PAPC[70,76] show an increased adhesive strength when the cytoplasmic tails were removed. These observations show that the mechanisms involved in homophilic interactions are different for the various δPcdhs and have been not fully elucidated yet.

Heterophilic cis interactions were also reported for δPcdhs and there is increasing evidence for an association between the δPcdhs and classical cadherins. Several studies have shown this association for different members of the δPcdh family and in different systems. PAPC, for instance, antagonizes the adhesion by C-cadherin[78] and even if a physical interaction has not been demonstrated, the formation of a complex between PAPC, C-cadherin and a leucine-rich repeat protein, Flrt3, has been

proposed in *Xenopus*[79]. δ2Pcdhs have been shown to associate with N-cadherin to form adhesive cis complexes[49]. Pcdh8/Arcadlin interacts laterally with N-cadherin (N-Cadh) in cultured hippocampal neurons of the rat and this association occurs at their transmembrane domains[80]. Pcdh8 interaction induces N-cadh endocytosis through a pathway, which involves the TAO2β kinase and p38 MAPK, and causes the downregulation of the adhesive activity of N-Cadh.

N-cadh forms a cis complex as well with Pcdh19 in the developing zebrafish.[81] Pcdh19 is the main responsible for this physical interaction, which is not lost in case of a mutation inhibiting N-cadh homophilic binding[82] (**Figure 10A**). In a Pcdh19 knock down zebrafish embryo, a malformation of the anterior neural tube can be observed, like in N-cadh mutant embryos, suggesting a collaboration of the two molecules in brain morphogenesis[83]. A stronger ability of Pcdh19 to form homodimers has been observed when Pcdh19 is in complex with N-Cadh. So far, one could conclude that N-cadh acts as a co-factor in cis to facilitate cell-cell adhesion by Pcdh19 in trans. Since another δ2Pcdh, Pcdh17, shows the same behavior with respect to interaction with N-Cadh, this suggests a more general ability of N-cadh to form cis-complexes with different Pcdhs, whereas Pcdhs are the main mediators of adhesion specificity[82]. The interaction with N-cadh is occurring also for all the members of the δ1Pcdh subfamily in zebrafish[84]. A regulation of N-cadh assembly by Pcdh10 has been shown at the cell-cell contacts via the recruitment of NAP1 and WAVE1, but direct interaction of N-cadh with Pcdh10 has not been shown[85]. Taken together, those data show that Pcdhs can directly or indirectly interact with classical cadherins, influencing in this way their adhesive behavior.

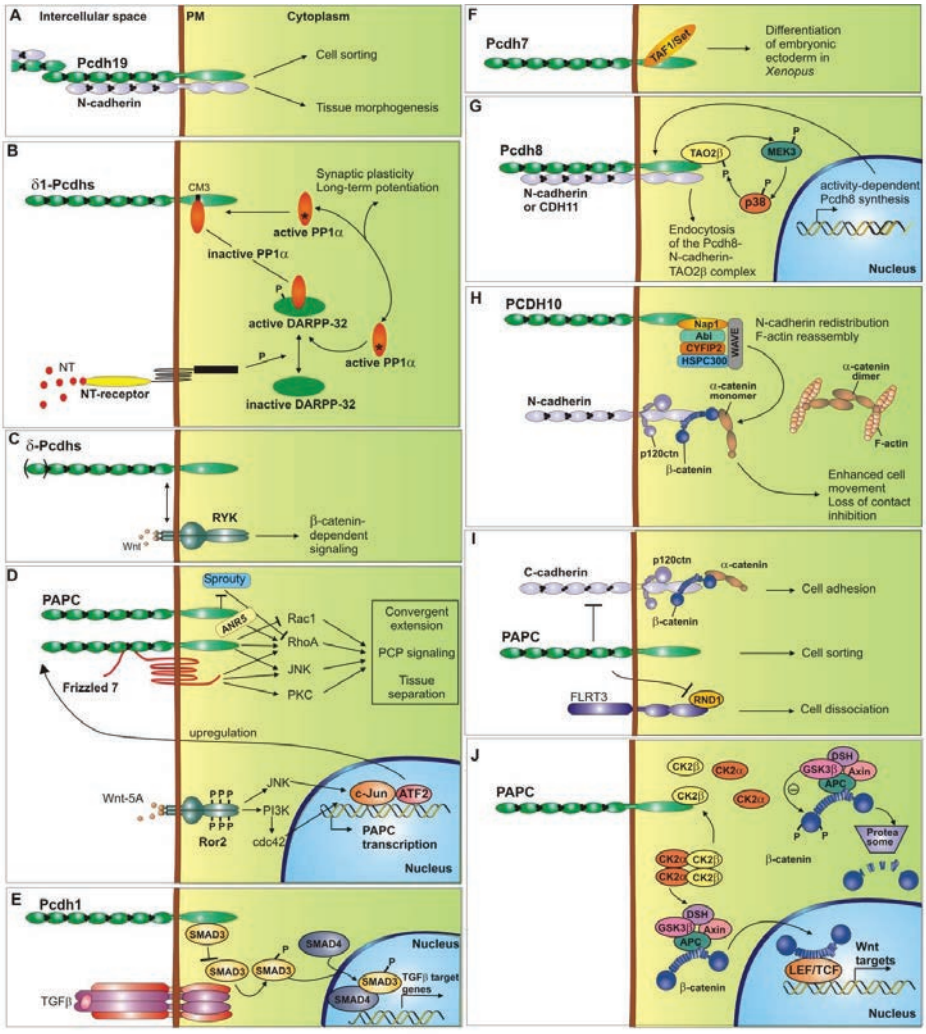


Figure 10 Schematic overview of δ-Protocadherins interactions

See text, I.3.4, for description references and acronyms. (A) Protocadherin 19 (Pcdh19) mediates a new mechanism of homophilic cell adhesion with N-cadherin acting as a required cis-cofactor. (B) δ1-protocadherins interact with PP1α, probably inactivating it. They might work synergistically with neurotransmitters (NT) to regulate synaptic plasticity. (C) Several δ-protocadherins were shown to associate with RYK, which functions as a transmembrane receptor for ligands of the Wnt family. Both δ-protocadherins and RYK were shown to be involved in the regulation of cell movements during convergent extension in zebrafish and frog, and of axon path finding in vivo *and* in vitro. *(D) Paraxial protocadherin (PAPC) interacts with Frizzled-7 to activate downstream signaling through RhoA, JNK and PKC, which leads to tissue separation during convergent extension movements in Xenopus embryos. Furthermore, PAPC interacts with ANR5 and Sprouty, which, respectively, activate or inhibit RhoA signaling, and thereby promote the planar cell polarity (PCP) pathway. PAPC expression is regulated by binding of Wnt-5A to Ror2, which in turn activates JNK signaling, leading to upregulation of PAPC. P, phosphotyrosine. (E) PCDH1 interacts with SMAD3 with subsequent inhibition of its*

TGFbeta dependent activation. This results in a suppression of TGFbeta induced gene transcription. (F) Protocadherin-7 (Pcdh7) interacts with the histone-regulating protein TAF1/Set; this interaction could be causally linked to differentiation of Xenopus embryonic ectoderm. (G) On neural activation, N-cadherin and cadherin-11 associate with protocadherin-8 (Pcdh8), which binds to the MAPKKK TAO2β. TAO2β activates p38 MAPK, which in turn feeds back to TAO2β, in the end triggering endocytosis of the Pcdh8-N-cadherin-TAO2β complex. (H) Protocadherin-10 (PCDH10) associates with Nap1, which is a known mediator of WAVE-mediated actin assembly. By recruiting Nap1 to cell–cell contact sites, PCDH10 might alter F-actin organization, which in turn causes N-cadherin redistribution. As the redistributed N-cadherin molecules can no longer induce contact inhibition, the result is uncooperative, accelerated cell movement. It is also shown the classical cadherin-catenin complex, and the binding of α-catenin dimers to F-actin. (I) PAPC influences morphogenetic movements and cell sorting in Xenopus embryos by down-regulating the adhesive activity of C-cadherin. Moreover, PAPC and C-cadherin form a functional complex with FLRT3. Unregulated FLRT3 causes excessive cell dissociation, but PAPC binding inhibits the ability of FLRT3 to recruit the small GTPase RND1, and this reduces the magnitude of FLRT3-induced cell dissociation. (J) PAPC can inhibit canonical Wnt signaling by interacting with CK2β. This interferes with CK2 complex formation, which leads to proteasomal degradation of β-catenin and subsequent inhibition of Wnt target gene transcription. (Figure modified after [62]).

I.3.4 Molecular Interaction Partners and Physiological functions

In fact, in view of the weak homophilic binding ability of δPcdhs, it has been speculated that intracellular involvement in signaling pathways might be the main role of the δPcdh family members, and evidence for this has indeed been reported in different studies. Also in support of this theory is the peculiar structure of the cytoplasmic domains (CD) of δPcdhs. As already mentioned, the CD of δPcdh is quite different form the one of classical cadherins, and neither the conserved p120 nor the conserved β-catenin binding cytoplasmic domains are present. Comparing the CD of the members of the family, it becomes clear that they are different not only from the one of classical cadherins but also between each other: their homology ranges from low to moderate[2,29].This might allow the members of the family to be able to interact with a relevant variety of different intracellular interaction partners.

The δ1 and δ2Pcdhs are expressed as both short and long isoforms, generated by alternative splicing and differing from each other by the size of their cytoplasmic domain and the presence or not of conserved motifs (CMs). All the long isoforms show CMs (CM1 and CM2 for δ1 and δ2Pcdhs; plus CM3 for δ1Pcdhs only) in their cytoplasmic domains[22] (**Figure 11**). This, together with a C-terminal PDZ-domain binding site[86], is what indicates that δPcdhs likely share some cytoplasmic interaction partners. A new conserved region has been recently identified in PAPC, the so-called DSR domain: it is part of the CM1 domain and is rich in aspartate (D) and serine (S) residues. Regulation occurs via phosphorylation of DRS by GSK3β, followed by polyubiquitination by the E3 ubiquitin ligase, β-TrCP. Similar mechanisms of regulation might occur for NC-Pcdhs as well but have not been shown yet[87].

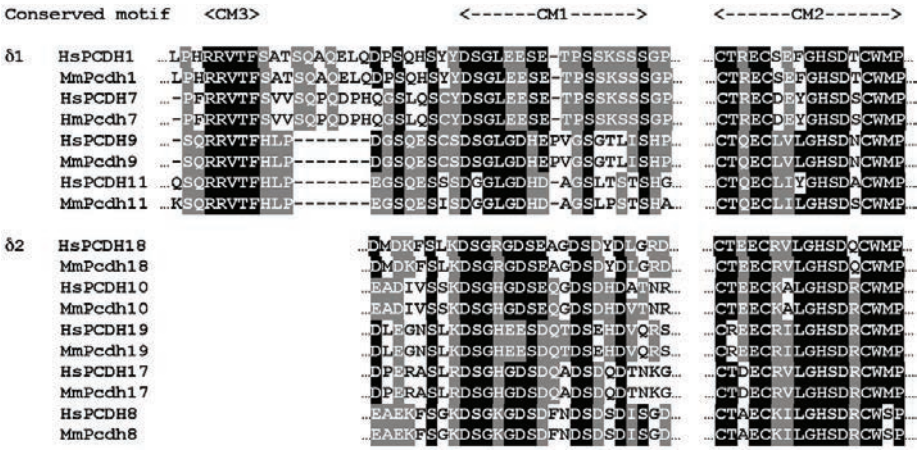


Figure 11 Multiple amino acid alignment of selected parts of the cytoplasmic domains of the δ-protocadherins expressed as long isoforms.
Sequences are human (Hs) or mouse (Mm). CM, conserved motifs (1 to 3). Upstream and downstream sequences are not shown to reduce the complexity of the figure. From[22].

In recent years different studies have focused on the identification of molecular interaction partners of δPcdhs (**Table 2**). Unfortunately, the binding molecules of the CMs remain largely elusive; only CM3, common for all δ1Pcdhs, is known to interact with PP1α[22,88] (**Figure 10B**).

Table 2 Summary of δPcdhs known interaction partners
References and description can be found in the text

	Gene symbol	Interaction partners
δ1	PCDH1	RYK, PP1α, SMAD3
δ1	PCDH7	PP1α, RYK,TAF1/Set
δ2	PCDH8	TAO2β, N-cadherin
δ2	PAPC	Frizzled-7, Sprouty, ANR5, CK2β, C-cadherin, FLRT3
δ1	PCDH9	PP1α, RYK
δ2	PCDH10	Nap1, RYK, DYNLT1
δ1	PCDH11X	PP1α, DYNLT1
δ1	PCDH11Y	PP1α, β-catenin
δ2	PCDH17	N-cadherin
δ2	PCDH18	Dab1, RYK
δ2	PCDH19	Nap1, RYK, N-cadherin
δ0	PCDH20	RYK

I.3.4.1 δ1Pcdhs and PP1α

Protein phosphatase 1 α (PP1α) is one of the major Ser/Thr phosphatases in mammalian cells. It is one of the 4 mammalian isoforms of PP1: PP1α, PP1β/δ, PP1γ1 and PP1γ2. PP1 regulates several processes such as glycogen metabolism, cell division, cell differentiation, cell death, signal transduction and pre-mRNA splicing in mammalian cells[89]. PP1α is mainly expressed in the nervous system, specifically in the brain. Its expression is upregulated in medium sized spiny neurons of the striatum where the regulation of DARPP-32 phosphorylation occurs. DARPP-32 stands for Dopamine and cAMP Regulated Phosphoprotein, 32 kDa. If phosphorylated, DARPP-32 inhibits the activity of PP1α. The latter has been implicated in regulation of long term synaptic plasticity in neurons and plays a role in memory formation and learning processes. PP1 isoforms are found in abundant levels in the cytosol and in the nucleus. PP1γ and PP1β translocate to the nucleus by binding to a co-transporter that contains a nuclear localization signal (NLS)[88,90,91]. Recent findings have shown that 14-3-3ζ binds to PP1α and causes its retention in the cytosol which suggests that 14-3-3ζ regulates nuclear trafficking of PP1α in mammalian cells[89]. Yoshida and colleagues isolated PP1α in a yeast two-hybrid screen for intracellular interaction partners of PCDH7[88] (**Figure 10B**). This interaction inhibits the enzymatic activity of PP1α on glycogen phosphorylase and it occurs via the RRVTF interaction motif or CM3 present in all δ1Pcdhs. As expected, follow-up studies have revealed that PP1α is a common interaction partner for all δ1-Pcdhs[22]. Since both PCDH7 and DARPP-32 inhibit the activity of PP1α, a synergistic collaboration between these two proteins has been proposed[88,92]. So far, there is no evidence published for other common interaction partners binding any of the CM domains of δPcdhs.

I.3.4.2 δPcdhs and RYK

δPcdhs and several γ–Pcdhs were isolated in a screening for receptor-like tyrosine kinase (RYK) associated proteins (**Figure 10C**). RYK is defined as an atypical member of the family of growth factor receptor protein tyrosine kinases and it is composed of a extracellular Wnt inhibitory factor domain, an intracellular atypical domain and a PDZ binding motif[93]. It has been defined atypical because it differs from other family members in a number of conserved residues, localized in the activation and nucleotide binding domains. It has been proposed that the recruitment of signaling proteins by RYK may occur to mediate the biological activities. Among those activities there is the stimulation of the Wnt signaling pathway and a clear involvement in the regulation of cell movement during convergent extension of zebrafish and frog[94,95] and in regulation of axon pathfinding in multiple

species. Notably, δPcdhs are required for the same phenomena. Furthermore, there is evidence for connection between the δPcdh family and the Wnt-triggered RYK signaling pathway. For instance PAPC (homologue of Pcdh8 in *Xenopus*) binds the Frizzled-7 receptor[96] (**Figure 10D**), and it is known that RYK binds Frizzled proteins and that PAPC is regulated via Wnt-5a[97]. Those are all indications of a connection between RYK and δPcdhs but a direct interaction has not been reported.

I.3.4.3 Pcdh1 and SMAD3

Besides the possibility of sharing interaction partners, each δPcdh may and does interact in a specific way with a different list of proteins, and this can be ascribed to differences in the other sequences of the CD besides the CMs. Pcdh1, for instance, interacts with SMAD3 in yeast two-hybrid (Y2H) overexpression assays[98] (**Figure 10E**). SMAD3 forms a SMAD3/SMAD4/JUN/FOS complex at the AP-1/SMAD site to regulate TGF-beta-mediated transcription and it is implicated in a list of functions including epithelial to mesenchymal transition (EMT) and epithelial repair. PCDH1 is mainly expressed in airway epithelial cells; its expression is strongly increased during mucociliary differentiation and this has been related to asthma[58,99]. Recently, an interaction of Pcdh1 with Smad3 has been confirmed via CoIP in 16HBE cell lysates[99]. Studies involving either overexpression of PCDH1 or its expression at the endogenous level have been carried out in HEK293T cells and bronchial epithelial cells, respectively, and demonstrated an inhibition of the TGF-β dependent activation of SMAD3 due to the interaction with PCDH1. If PCDH1 is knocked down in primary human bronchial epithelial cells, fibronectin expression was induced. Interestingly, increased deposition of fibronectin (and collagen) into the subepithelial space of the airways is observed in all forms of asthma. The Pcdh1-Smad3 interaction is thus able to suppress the activity of SMAD3, suggesting a collaboration between two asthma-related genes[99]. However, the detailed underlying mechanisms have not been elucidated yet.

I.3.4.4 Pcdh7 and TAF1/Set

Pcdh7 interacts also with TAF1/set besides with PP1α and RYK[100](**Figure 10F**). Pcdh7/NFPC has four isoforms (7a, 7b, 7c and 7c1), and PCDH7c and 7c1 have CM1, CM2 and CM3 motifs. However, all PCDH7 isoforms interact with TAF1. Disruption of the ectodermal integrity in the *Xenopus* embryo due to the ectopic expression of a dominant-negative form of Pcdh7 lacking the ectodomain can be rescued by the expression of TAF1[75,100]. Furthermore, NFPC and TAF1 are responsible for cell-cell adhesion in the neural ectoderm: disruptions in either NFPC or TAF1 result in a failure of the neural tube to close because of the sudden lack of organization of the cells in the dorsal neural fold.

Accordingly, the two proteins are necessary for the proper apical localization of F-actin in neural fold cells[101]. Both proteins are also expressed in retinal ganglion cells (RGCs) of *Xenopus* [102]: the inhibition of either one of the two interferes with proper RGC development. NFPC is not only expressed in RGCs but also within the retina, on RGC axons and by cells within the optical tract and tectum. A recent study has shown that NFPC contributes to axon behavior at multiple points in the optic pathway, therefore playing an important role in RGC axonogenesis[103].

I.3.4.5 Pcdh8 and TAO2β

Pcdh8, which is named Arcadlin in rats, has been shown to interact with N-cadh at the synaptic membrane of rat hippocampal neurons[71,80] (**Figure 10G**). Rat Pcdh8 possesses homophilic binding affinity. In *Xenopus* and zebrafish, arcadlin/PAPC plays an important role in homotypic cell-cell adhesion and in gastrulation movements of the paraxial mesoderm[76]. Homophilic interactions in trans between arcadlin extracellular domains trigger a signal transduction cascade connected to p38 MAPK: the intracellular domain of PCDH8/arcadlin interacts with protein kinase 2β (TAO2β), which in turn activates the p38 MAPK pathway and subsequently promotes endocytosis of N-cadherin. N-cadherin regulates spine dynamics and maintains the shape and density of spines. The PCDH8-TAO2-p38 MAPK pathway mediated endocytosis of N-cadherin may therefore result in the disruption of synaptic adherent junctions[71,80].

I.3.4.6 Pcdhs and NAP1

Several studies have suggested that PCDHs participate in neural circuit assembly although their respective specific roles in neuronal development remain largely unknown. In the light of this need for more data to understand roles and mechanisms, an intriguing interaction partner of δ-PCDHs is Nck-associated protein 1, Nap1[85] (**Figure 10H**). Nap1 is a core protein of the WAVE complex, a main regulator of actin which plays a central role in many cellular processes including adhesion, migration, division and fusion. The different roles played by the WAVE complex in animals goes from embryogenesis to cancer invasion and metastasis, including neuronal morphogenesis and plasticity[104,105]. Nap1 modulates actin nucleation, by forming a pentameric complex with WAVE, PIR121, Abi1/2, and HSPC300[106]. It is expressed in the developing cortical plate where it plays an important role in cortical laminar-specific neuronal differentiation. This involves a major contribution to the neuronal connectivity in the cerebral cortex: if Nap1 is lost the neuronal differentiation is disrupted[106]. PCDH10 has been shown to interact with Nap1 and to associate with WAVE1 and Abi in pull-down experiments[85]. If PCDH10 is overexpressed, Nap1 and WAVE1 are

recruited to the cell-cell contacts, which causes a reorganization of F-actin and N-cadh with consequent accelerated cell migration. In this way, N-cadh loses the ability of inducing contact inhibition so that cells start to behave in uncooperative manners. Though the expression of PCDH10 does not alter the migration of single cells, groups of cells are negatively influenced. To conclude, the PCDH10-Nap1 interaction promotes contact-dependent cell motility, indicating a dependence on homophilic interactions between neighboring cells[85,107]. Chicken Pcdh19 was also shown to interact with Nap1 (and interactions of Nap1 with Pcdh1 and Pcdh9 were likewise suggested)[77]. Despite the presence of the CMs in the CD of chicken Pcdh19, the interaction was suggested to occur in a domain located at the C-terminus. A sequence alignment of conserved regions of δPcdh cytoplasmic domains (of mouse OL-Pcdh, mouse Pcdh18, mouse Pcdh17, human Pcdh11, chicken Pcdh9, mouse Pcdh8, mouse Pcdh7, and chicken Pcdh1) showed the presence of a very short yet conserved region at the C-terminus, which was called CM4 and was proposed to be the interaction site of Nap1[85]. A smaller conserved nine-residue sequence, RSFSTFGKE, within the CM4 element, has been recently described to be responsible for the Nap1 binding[107]. In this same study it has been proposed that the interaction with Nap1 is indirect and requires the formation of the whole pentameric WAVE complex[108]. Recently, Nap1 has been shown to interact with the intracellular domain of zebrafish Pcdh18b and in case one of the two proteins was depleted, very similar effects on axon branching and dynamics were observed. As described for the PCDH10-Nap1 interaction, loss of either Pcdh18b or Nap1 reduces the density of filopodia and results in reduced axon arborisation, suggesting a dependency on interactions between neighboring cells[109]. To conclude, Nap1 and the WAVE complex may be general downstream effectors of δPcdhs, and the δPcdhs appear to be regulators of the WAVE complex and actin dynamics during nervous system development.

I.3.4.7 Pcdh18 and DAB1

Pcdh18 also interacts with mouse Disabled homolog 1 (mDab1). Dab1 is a key regulator of Reelin signaling and mediates neural circuit formation during brain development: mutations in Dab1 cause ataxia, cerebellar hypoplasia and disorganization of the laminar structure of the CNS[110–113]. Tyrosine-phosphorylated mDab1 associates with the SH2 domains of Src, Fyn, and Abl[110]. The interaction with Pcdh18 was shown in a yeast two-hybrid screening and occurs at the C-terminal 243 residues of mPcdh18, in the conserved motif NPTS, unique for Pcdh18[114]. Fyn, like Dab1, is a downstream mediator of the Reelin pathway: mice defective for Reelin, Dab1 or Fyn show very similar phenotypes characterized by inverted cortical layers. It has been shown that PCDH-α/CNR interacts with Fyn, suggesting that Pcdh18 might be involved in Reelin dependent signaling in view of

its direct interaction with Dab1. Furthermore, PAPC interacts with Sprouty (**Figure 10D**), a protein with inhibitory activity on receptor tyrosine kinase signaling proteins. The interaction with PAPC inhibits the activity of Sprouty during convergent extension movements in Xenopus[115]. ANR5 (ankyrin repeat protein 5), a protein that regulates cell movements during gastrulation, is also an interaction partner of PAPC.[116] These interaction partners of different NC-Pcdhs suggest an important role played by NC-Pcdhs in cell movement and motility.

I.3.4.8 Pcdhs and tissue morphogenesis

Identification of putative interaction partners is a helpful starting point to understand the physiological functions of the δPcdh family. As of now it becomes clear from the description of the already known interaction partners that δPcdhs are involved in tissue morphogenesis and brain development. Some examples have already been mentioned in this section but more studies need to be conducted with the aim of identifying the functions of δPcdhs. The regulation of tissue morphogenesis has been studied in different organisms including *Xenopus*, chicken and mouse, and different δPcdh members have been involved in different ways in this important regulation. *Xenopus* PAPC, for instance, is one of the best studied δPcdhs and has been shown to regulate both canonical and non-canonical Wnt signaling. The casein kinase 2 (CK2) complex, which is involved in the regulation of Wnt target genes, activation of dishevelled and stabilization of β-catenin in the cytoplasm, has been shown to interact with PAPC[117] (**Figure 10I and 10L**). This interaction antagonizes the cytoplasmic stabilization of β-catenin with subsequent inhibition of Wnt target gene transcription[117]. But PAPC interacts with the Wnt receptor Frizzled-7 as well, interfering in this way with the downstream activation of the planar cell polarity (PCP) pathway. Moreover, as already mentioned, PAPC interacts with ANR5 and Sprouty, which are involved in the regulation of RhoA signaling, promoting the PCP pathway[96,115,116,118]. Furthermore, the JNK-dependent activation of Wnt-5A-Ror2 is responsible for the regulation of PAPC[118] (further information in **Figure 10D** and reviewed in[62]). Recent findings have shown that Protein Tyrosine Kinase 7 (PTK7), a tyrosine kinase receptor lacking catalytic activity, binds Ror2 to regulate PAPC expression and is required for PAPC activation as induced by Wnt5a: the intracellular domain of PTK7 is released after Wnt5a stimulation and migrates to the nucleus to activate PAPC transcription[119]. Surprisingly, Pcdh8-null mice do not show any obvious phenotype[120,121]. Since an overlapping function with *Xenopus* PAPC has been shown for Pcdh10 in zebrafish, a functional redundancy for Pcdh8 and Pcdh10 has been proposed. As already mentioned, PAPC can downregulate the adhesion activity of the classical C-cadherin, which induces a change in morphogenetic movements[79]. Recently, a new interaction partner for PCDH10 and

PCDH11X has been identified (U. Brunner and I. Kahr, not published): DYNLT1. Dynein light chain or Tctex type 1 (DYNLT1) seems to interact with different regions of these two PCDHs. Further studies are required to investigate the function of this interaction.

I.3.4.9 Pcdhs and brain development

Concerning brain development we can conclude that δPcdhs appear to be involved in different functions including the maintenance of the complexity of the neural network due to combinatorial expression, in particular of Pcdh-γs. Moreover, δPcdhs are involved in axonogenesis, dendritogenesis and axon guidance as different studies have shown, and they interact with a number of proteins involved in neuronal development, hence possibly modifying their activity. Despite these findings, underlying molecular mechanisms have still to be elucidated. For example, in a recent study Compagnucci and colleagues[122], tried to elucidate the function of PCDH19 in human brain development: they used induced pluripotent stem cells (iPSCs), able to mimic the closure of the neural tube, in order to characterize the location and timing of expression of PCDH19 during cortical neuronal differentiation. Generation of the neural tube was resembled by the formation of "neural rosettes", which are composed of Neuronal Stem Cells (NSC) arranged radially around a lumen. The study focused in particular on the establishment of apical-basal polarity and the geometry of proliferation. PCDH19 localized in the center of the neural rosette, which corresponds to the lumen of the neural tube, suggesting a role in the development of the human brain architecture. The same study showed PCDH19 localization at the spindle pole of the dividing cells. This may indicate a role of PCDH19 in the control of asymmetric versus symmetric cell division during *in vitro* neurogenesi[122]. As I mentioned above, many more data and studies are needed to elucidate the mechanisms involved in Pcdh physiological functions.

I.3.5 Human diseases and δPcdhs

δPcdhs are expressed at high levels in the brain. So, it is not that surprising that many of them are involved in neurological disorders, such as autism, schizophrenia, epilepsy and Alzheimer's disease. A role in neurological disorders has been reported for PCDH7[123], PCDH10[124] , Pcdh8 and Pcdh17[125], PCDH11X and PCDH11Y[126,127] and PCDH18[128]. The most convincing evidence for involvement in neurological disorders comes from studies on *PCDH19*, which is the most relevant gene in epilepsy after SCN1A[129]. PCDHs have also been shown to be involved in other diseases and symptoms, for example, asthma and bronchial hyper-responsiveness (BHR)[55]. Recent reports also describe a role for δPcdhs in the immune system[130–133]. Mounting evidence suggests that δPCDHs can function as either

tumor suppressors or proto-oncogenes in various non-neuronal human tumors (reviewed in van Roy, 2014)[11](**Table 3**). For instance, analysis of pancreatic cancers showed missense mutations of PCDH9, -17, and -18[134]. Strong evidence was reported for frequent epigenetic inactivation of the *PCDH10* gene in various human cancers but not in matched normal tissues[61,135,136]. Furthermore, loss of expression of PCDH1, PCDH9, PCDH10, PCDH17 and PCDH20 has been shown to correlate with poor prognosis or resistance to treatments[137–142].

In the next paragraph we will describe the role of δPCDHs in cancer. Both direct and indirect evidence indicates that δPCDHs might play important roles during tumorigenesis, either as tumor-suppressor genes or as oncogenes. The roles played by δPcdhs in neurological diseases will be reviewed following the cancer section.

Table 3 δ-Protocadherins in cancer: an overview
*(AIG, anchorage-independent growth; APO, decreased apoptosis induction; AUT, decreased autophagy induction; CAM, positive in chick chorioallantoic membrane assay for neoangiogenesis; CCA, decreased cell cycle arrest; MCF, positive in monolayer colony formation (clonogenicity); MIA, positive in Matrigel invasion assay; MPA, positive in monolayer proliferation assay (adherent cells in vitro); MWH, positive in monolayer wound healing (cell migration); PCDH, protocadherin; PCDH11X, X-linked protocadherin 11; PCDH11Y, Y-linked protocadherin 11; TSG, tumor-suppressor gene; XMC, xenograft metastatic colony formation (in athymic mice); XTG, xenograft tumor growth (in athymic mice). *The tumor-specific changes listed are only those reported for human tumor samples rather than for tumor cell lines. Data on somatic mutations (mostly substitution missense) were obtained from the Sanger Institute Catalogue of Somatic Mutations in Cancer website and from the cBioPortal for Cancer Genomics, and the data are given for those PCDHs and tissue types for which the percentage of samples with non-synonymous mutations exceeds 7% and more than ten cases were analyzed. Human-specific paralogue, expressing an Y-encoded cytoplasmic isoform of PCDH11, also known as PCDH-PC). (Table from[11])*

PCDH type	Tumor type*	Reported tumor-specific changes*	Activity	Refs
PCDH1	Medulloblastoma	Decreased expression	Shorter survival	142
	Pancreatic	Epigenetic silencing	TSG?	143
PCDH7	Breast	Increased expression in brain metastatic tumor cell populations	Metastasis gene?	144
	Breast	Upregulated in bone metastatic primary tumors	MPA, MWH, MIA and XMC	145
PCDH8	Colorectal, mantle cell lymphoma, pancreatic, renal cell and nasopharyngeal	Epigenetic silencing	TSG? MCF and MWH	146–149
	Breast	Epigenetic silencing and/ or somatic mutations	AIG and MWH	150
	Prostate	Homozygous deletion‡	TSG?	151
PCDH9	Glioblastoma multiforme	Decreased expression	TSG in glioblastoma multiforme? Higher	152, 140,15

	and other gliomas		tumor grade, shorter survival, APO, CCA and AIG in glioma	[3]
	Colorectal, endometrial, esophageal and prostate	Somatic mutations‡ (mostly substitution missense)	TSG?	[154,151]
PCDH10	Breast, gastric, colorectal, pancreatic, cervical, testicular germ cell tumors, various lymphoma types, multiple myeloma, B cell and T cell acute lymphoblastic leukemia and reactive follicular hyperplasia	Epigenetic silencing	TSG? For multiple myeloma: MCF, CCA and CAM. For acute lymphoblastic leukemia: chemotherapy resistance	[143,135,136,141,155–158]
	Nasopharyngeal, esophageal, breast, gastric, hepatocellular, non-small-cell lung	Epigenetic silencing	TSG? MCF, AIG, MWH and MIA. In addition, for gastric: shorter survival, MPA, APO and XTG	[61,139,159]
	Medulloblastoma	Epigenetic silencing and/or homozygous deletion	TSG? MWH	[160]
	Lung, gastric and colorectal	Somatic mutations‡ (mostly substitution missense)	TSG?	[154,151]
PCDH11X	Prostate	Missense mutations	Lethal metastatic tumors	[161]
	Lung, esophageal, gastric, as well as head and neck	Somatic mutations‡ (mostly substitution missense)	TSG?	[154,151]
PCDH11Y||	Prostate	Upregulated	Androgen-independent growth, anti-apoptotic, neuroendocrine trans differentiation and resistance to cytotoxic drugs	[162]
PCDH17	Pancreatic and gastric	Epigenetic silencing	TSG?	[163]
	Esophageal squamous cell	Epigenetic silencing and/or homozygous deletion	Poorer differentiation status of tumors; MPA, CCA, MWH and MIA	[138]
	Gastric and colorectal	Epigenetic silencing and/or homozygous deletion	TSG? MCF, AIG, APO, AUT and XTG	[164]
	Colorectal, esophageal, gastric, lung, skin cutaneous melanoma and prostate	Somatic mutations‡ (mostly substitution missense)	TSG?	[154,151]
PCDH18	Skin cutaneous melanoma, esophageal and gastric	Somatic mutations‡ (mostly substitution missense)	TSG?	[154,151]

PCDH19	Gastric and endometrial	Somatic mutations‡ (mostly substitution missense)	TSG?	[154] [151],
PCDH20	Non-small-cell lung	Epigenetic silencing	Shorter survival, MCF and AIG	[137]
	Prostate and skin cutaneous melanoma	Homozygous deletion‡	TSG?	[151]

I.3.5.1 δPcdhs in Cancer

PCDH8 is a strong candidate tumor suppressor gene (TSG) in breast cancer[150]. In the recent years early screening supporting early diagnosis has decreased the incidence of this kind of cancer. Despite that progress, breast cancer is still the first cause of death for woman in the U.S.: in 2015, over 40,000 women died from breast cancer[165]. PCDH8 has been shown to be inactivated through either mutation or epigenetic silencing in a high fraction of breast carcinomas. The lack of PCDH8 expression associates with loss of heterozygosity and partial promoter methylation[150]. The same study showed that PCDH8 inhibited the proliferation of the mutant tumor cell line HCC2218 as well as the migration of the mammary cell line MCF10A. PCDH8 has been proposed to be a TSG also in nasopharyngeal carcinoma (NPC),[149] an epithelial malignancy with an unusual geographical and ethnic distribution and with a multi-step pathogenesis, which depends on Epstein–Barr virus infection, dietary and environmental factors, and accumulation of genetic and epigenetic alterations of multiple cancer genes. Promoter methylation is the key mechanism for the inactivation of PCDH8 expression in NPC. The *PCDH8* gene is found to be silenced also in colorectal cancer[146], mantle cell lymphoma[147], and renal cell carcinoma[166]: mutation and/or promoter methylation have been shown and exogenous expression of PCDH8 suppress proliferation in the mutant tumor cell line HCC2218 and migration of the mammary cell line MCF10A[150].

Mounting evidence points at a role of PCDH10 as a tumor suppressor. Epigenetic silencing by *PCDH10* promoter methylation has been demonstrated in different kinds of cancer, including breast cancer[135], lymphomas[136], cervical cancer[155], testicular cancer[156], prostate cancer[158], multiple myeloma[157], hepatocellular carcinoma[167] and endometrioid endometrial carcinoma[168]. Ectopic PCDH10 expression in nasopharyngeal and esophageal carcinoma cell lines reduced clonogenicity, anchorage-independent growth, migration potential, and *in vitro* invasion into Matrigel[61]. Recently, it has been shown that *PCDH10* is a candidate tumor suppressor gene in medulloblastoma[160]. Medulloblastoma is an aggressive, fast-growing and high-grade tumor: It is relatively rare but is the

most common cancer type in children. It is an embryonal tumor, which arises from cells at the earliest stage of development in the cerebellum.

For gastric cancer, *PCDH10* promoter methylation was detected at early stages of carcinogenesis and was associated with poor prognosis[139]. Re-expression of PCDH10 in gastric cancer cells inhibited tumor growth in nude mice, induced cell apoptosis, and inhibited cell invasion. Inactivation of *PCDH10* by promoter methylation is a frequent pathogenic event in multiple myeloma (MM): a recent study showed that PCDH10 antagonized cell proliferation in multiple myeloma via downregulation of Wnt/β-catenin/BCL-9 signaling[169]. Furthermore, PCDH10 induced apoptosis in MM cells which is mediated by activation of caspase-3 and PARP and by downregulation of the anti-apoptotic proteins Mcl-1, Bcl-2, Bcl-xL, survivin, XIAP and cIAP-1[170]. The downregulation of these proteins may be due to the inhibition of the NF-κB pathway. This suggests that inhibition of the constitutively active NF-kB pathway in MM might be the mechanism of the pro-apoptotic function of PCDH10[170]. Moreover, a recent study has shown that PCDH10 is a functional tumor suppressor in colorectal tumorigenesis and that it plays a pivotal role in restraining liver metastasis[171]. Finally, silencing of *PCDH10* contributes to chemotherapy resistance of leukemia, and the methylation status of *PCD10* can be used as a diagnostic test for cervical carcinogenesis on the basis of cervical scraping[172,173].

PCDH10 plays a role in the apoptosis mechanisms also in gastric cancer cells: restoration of the PCDH10 expression in the MKN45 gastric cell line induces cell apoptosis via upregulation of Fas, caspase-8, Jun, and CDKN1A[139]. Also PCDH17 has been shown to be involved in apoptosis regulation: expression of PCDH17 induces apoptosis in gastric and colorectal cancer cells[164]. *PCDH17* is a tumor suppressor gene as well for esophageal squamous cell carcinoma[138], laryngeal squamous cell carcinoma[174], gastric cancer cell lines[163], and colorectal carcinoma[164]. Recent findings in colorectal carcinoma suggested that AMPK (AMP-activated protein kinase) may be involved in PCDH17 expression regulation: Se-allylselenocysteine (ASC)-induced autophagic cell death is enhanced by upregulation of phosphorylated AMPK and demethylation of the *PCDH17* promoter, which was suppressed by AMPK inhibitor, showing that activation of AMPK signaling plays a role in the PCDH10 expression level[175].

The activation of the EGFR/MEK/ERK signaling pathway through loss of PCDH17 promotes metastasis in hepatocellular carcinoma (HCC): a recent study has shown that PCDH17 expression correlates with overall prognosis and metastasis *in vivo.* Furthermore, PCDH17 has been shown to inhibit metastasis via EGFR/MEK/ERK signaling pathway in an experiment using a panel of HCC cell lines[176] .

So far, there is only weak evidence for the role of PCDH20 and δ1-PCDHs in cancer: the *PCDH20* promoter is frequently silenced in non-small cell lung cancer with poor prognosis, and upregulated expression of PCDH1 in medulloblastoma patients predicted survival[142]. *PCDH9* is a candidate TSG in glioblastoma and the loss of its expression correlates with poor prognosis. Hence, PCDH9 has been proposed as a prognostic biomarker (further examples in **Table 3**).

As already mentioned, despite the many studies and lines of evidence about the role of δPCDHs as TSG, some members of the family can behave also as proto-oncogene, although much less has been reported on this possibility. So, different studies have reported PCDH7 to have an oncogenic activity. Bos and colleagues showed a strong upregulation of PCDH7 expression in brain metastasis of breast cancer[144] .A strong upregulation has been shown also in case of medulloblastoma[177], in androgen-independent prostate cancer cells[178], and in melanoma with poor prognosis[179]. Furthermore, inhibition of PCDH7 suppresses cell proliferation, migration, and invasion in breast cancer cells and PCDH7 has been found to be upregulated in bone metastatic breast cancer tissues[145]. PCDH11X has been picked up in a screening to identify protein altering point mutations in lethal prostate cancer[161], and PCDH11Y negatively influences apoptotic sensitivity of prostate cancer cells and it is enriched in advanced castration-resistant prostate cancers[162]. Gene expression profile of B-cell chronic lymphocytic leukemia (CLL) identified *PCDH8* as a downregulated gene in patients with poor outcome indicating an oncogenic function of PCDH8 in leukemia[180].

I.3.5.2 δPcdh in Neurological Disorders

Different studies have shown that δPcdh are expressed in the brain and that they can play important roles in brain development and in the regulation of brain functions. Given these characteristics, an important role of δPcdh in different neurological disorders can be expected.

I.3.5.2.1 The Autism spectrum disorders (ASDs) and the Rett Syndrome

Autism spectrum disorders (ASDs) are conditions of coexistence of autism, Asperger syndrome and pervasive developmental disorder not otherwise specified (PDD-NOS) with epilepsy, intellectual disability or bipolar disorders: this group of childhood diseases is characterized by a prominent impairment in social functioning, specifically in social communication, with the presence of stereotypic and repetitive behaviors with restricted interests[181,182]. A twin study conducted in California[183] has shown that susceptibility to ASD has a moderate genetic heritability component together with a substantial environmental component. By association studies, which are sensitive to

allelic heterogeneity, a list of 190 genes (including *PCDH9* and *PCDH10*) in association with ASDs was proposed[108,184]. The confirmed list was divided into two parts: first, a group of genes or *loci* with high penetrance but mutated only in a limited number of individuals and second, the group of the so-called susceptibility genes to ASD. In these categories the genetic variations are different: for the first group, genetic variations are more often *de novo* or rare point mutations, copy number variations (CNVs), and cytogenetically detected deletions/duplications; while the genetic variations of the second group include SNPs or inherited CNVs observed in the general population and associating with low risk for ASD[185]. The genetic causes of ASD are thus multiple and diverse, although ASD-associated genes often play roles in the development and function of synapses. A recent study in *post mortem* ADS patients has clearly shown a reduction in the number of neurons in the amygdala, fusiform gyrus, and cerebellum, while the expression of genes involved in neuronal synapsis was downregulated in the brain[186]. Therefore, it has been speculated that ASDs are synaptic defect diseases[108].

Different studies indicate that several δPcdhs can be involved in cognitive dysfunction. For example, as mentioned above PCDH9 has been connected to ASDs in associative studies: recurrent structural variations such as copy number variation in *PCDH9* introns have been found in autistic patients. This association has been further demonstrated by the finding of downregulated *PCDH9* transcript levels in lymphoblasts of ASD patients[187–191]. These associations have been shown also in other species, for example in dogs[192]. Pcdh9 KO mice show a reduced thickness of the cortex, which is a symptom observed in several psychiatric disorders, including ASD. Although the molecular mechanism of this cortical phenotype is not understood, it suggests the role of PCDH19 in ASDs[192]. Several *loci* were mapped in a study with the aim of identifying inherited causes of ASD performed in families in which parents share ancestors, enhancing the role of inherited factors with autism. A large relevant homozygous deletion was found in a *PCDH* gene cluster close to *PCDH10* (on 4q28.3 between the *PCDH10* and the *PCDH18* locus), but evidence on a specific role for *PCDH10* is so far speculative[193]. Furthermore, PCDH10 is known to be required for cerebellar development and synapse formation of cerebellar Purkinje cells and for axon guidance in developing brain[194,195]. A dysfunction in these mechanisms can be related to the symptoms of ASDs[184,196]. Pcdh10 KO mice show defect in guidance of axons extensions in the ventral telencefalon[59], the brain part responsible for communication and behavior. The dysregulation of those functions is also compatible with ADS phenotype.

Furthermore, the autism syndrome has been shown to be a coexistent disturbance for patients with epilepsy due to *PCDH19* mutations (see below)[197]. Deregulation of PCDH19 in patients with PCDH19

related female epilepsy (PCDH19-FE) contributes to the occurrence of ADS causing the coexistence of the two phenotypes. Besides this, different missense variants of *PCDH19* gene, generally affecting the extracellular domain, have been identified in males showing ASD. A single missense variant, affecting the C-terminal of PCDH19, has also been identified in a boy with Asperger syndrome[198,199]. Taken together those studies indicate that PCDH19 is involved in ASD by different mechanisms.

Until the current edition of the *Diagnostic and Statistical Manual of Mental Disorders*, DSM-V, the Rett syndrome was included in the list of autism-related conditions, together with Asperger syndrome and childhood disintegrative disorder. Andreas Rett first classified this disorder in 1966[200]. He noticed a recurrent pattern in female patients such as regression after an early normal development, inability to use the hands with purpose and compulsive hand movements. The Rett condition was recognized after a study which included a series of 35 European girls with similar phenotypes[201]. Despite children with Rett syndrome exhibit behaviors typical of autism patients (for example repetitive hand movements, prolonged toe walking, body rocking, and sleeping problems), the Rett syndrome is not classified as an autistic syndrome. MeCP2 (methyl-CpG-binding protein 2) is a protein present in any kind of cell but highly expressed in brain where it is important for the function of several cell types, including neurons. The function of the protein is not completely clear but a role in the maintenance of neuronal synapses has been speculated on. Different mutations in the *MECP2 gene* have been shown to cause for example severe neonatal encephalopathy, PPM-X syndrome and Rett Syndrome. About 400 different MECP2 mutations have been identified and these can cause changes in the structure of the protein or in the amount of protein produced. *MEPC2*-null mice show a neurological phenotype similar to the one of Rett Syndrome patients indicating the role of the loss of MeCP2 function in the pathology of the syndrome. Thus, MeCP2 appears to be important for neuronal development and maturation, synaptic activity, learning and memory[123,202,203]. Miyake and colleagues[123] performed a genome microarray screen to reveal sites potentially regulated by MeCP2: within this list they showed that *PCDH7* and *PCDHB1* promoter activities were down-regulated by MeCP2. MeCP2 binds to the upstream regions of the protocadherin genes *PCDH7* and *PCDHB1* in human neuroblastoma SH-SY5Y cells, and PCDHB1 has been found to be upregulated in *post mortem* brains of Rett syndrome patients. This study indicates an association of dysregulation of protocadherins with the neuronal and synaptic defects typical of Rett patient's brains[123].

I.3.5.2.2 Schizophrenia and bipolar disorders

Schizophrenia (SZ), from Greek *skhizein*, to split + *phrēn*, "mind", is a long-term neurological disease. It is characterized by loss of normal thoughts, speech and behavior. It occurs with an incidence of approximately 0.5–1% with no gender differences. In 1887 the German psychiatrist Emil Kraepelin was the first to characterize the disease and he defined it "dementia praecox", making a clear difference with maniac depression disease. In 1911 the disorder was given the name we use nowadays, which indicates the fragmented thinking behavior. Different studies have revealed a long list of possible causes of schizophrenia. The environment plays an important role: for example, exposure to viruses or infection by the parasite Toxoplasma gondii[204], or malnutrition in the first stages of gestation. The use of addictive substances like mind-altering drugs or cannabis can be a risk factor too. Of course, interference with the normal functions of neurotransmitters such as dopamine and glutamate can cause schizophrenia as well. Schizophrenia has a strong genetic base[205]. Twin studies indicate a hereditability of around 80% but since the cause is never just one genetic variation, like for ASDs the identification of specific molecular genetic profiles for SZ is not straightforward[206]. The *PCDH8* gene maps in a region of chromosome 13q[207], where a linkage with SZ has been shown[208]. In a screen for polymorphisms in 30 schizophrenic patients, 9 *PCDH8* single-nucleotide variants have been found and proposed as a predisposition factor for SZ in families[209]. PCDH8 has been related to mental retardation in a study carried out on patients with retinoblastoma, dysmorphic features and developmental delay, and bearing a 13q14 deletion: this deletion affected both *PCDH8* and *PCDH17*[125]. Increased *PCDH17* mRNA expression has been found in the Brodmann's area 46 (BA46) of the brains from patients who have been showing SZ for less than 7 years[210]. Also *PCDH11X/Y* genes have been studied to understand their involvement in mental disorders. Durand and colleagues identified two PCDH11Y polymorphic amino acid changes in autistic patients, but the observation did not show statistically significant relevance when compared with controls[211]. A single splice mutation of PCDH11X has been reported in a single case of SZ,[198] together with a rare variant of PCDH19. More detailed functional analysis is still required to elucidate the mechanisms involved in the correlation between PCDHs and SZ.

I.3.5.2.3 Epilepsy

In epilepsy syndromes the normal pattern of neuronal activity is compromised. These syndromes influence emotions and behavior and, clinically, patients show seizures (disruption of the electrical communication between neurons), muscular spasm and loss of consciousness. Over 65 million

people suffer of epilepsy around the World and it can occur in any person and at any age. Different features characterize different epileptic syndromes, including the type of seizures, the age of symptomatic evidence, the causes and the severity of the seizures, genetic characteristics, and many more: the most common way to define the syndromes is by evaluating symptoms or by determining the part of the brain where the seizure starts. Epilepsy and mental retardation limited to females (EFMR) is an infantile onset epilepsy syndrome with or without intellectual disability (ID) and autism: several mutations of PCDH19 have been found in families with EMFR[129]. It is an X-linked disorder with an unusual inheritance pattern since it affects only carrier females with male sparing of the phenotype[212]. To investigate the role of PCDH19 in female patients with epilepsy, a group of 117 female patients with febrile seizures (FS) was studied: PCDH19 was mutated in 11% of cases showing focal epilepsy or Dravet syndrome-like phenotypes, with or without mental retardation and autistic features indicating that *PCDH19* defects play a major role in infantile-onset epilepsy in females[213]. Most of the mutations occur in the extracellular domain and only frameshift mutations have been shown in the intracellular domain of PCDH19. EMFR or PCDH19-FE has been shown to be very frequent in the families analyzed, identifying *PCDH19* to be the second most relevant gene in epilepsy after SCN1A. Over 60 different mutations in the *PCDH19* gene have meanwhile been identified suggesting that different mechanisms are involved in PCDH19-EF development[212,214].

I.4 References

1. Gumbiner, B. M. Regulation of cadherin-mediated adhesion in morphogenesis. *Nat. Rev. Mol. Cell Biol.* **6,** 622–34 (2005).

2. Halbleib, J. M. & Nelson, W. J. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Dev.* **20,** 3199–214 (2006).

3. Lien, W.-H., Klezovitch, O. & Vasioukhin, V. Cadherin-catenin proteins in vertebrate development. *Curr. Opin. Cell Biol.* **18,** 499–506 (2006).

4. Redies, C., Vanhalst, K. & Roy, F. van. δ-Protocadherins: unique structures and functions. *Cell. Mol. Life Sci.* **62,** 2840–2852 (2005).

5. Hulpiau, P. & van Roy, F. Molecular evolution of the cadherin superfamily. *Int. J. Biochem. Cell Biol.* **41,** 349–69 (2009).

6. van Roy, F. *The molecular biology of cadherins*. (Elsevier, 2013).

7. Priya, R. & Yap, A. S. Active tension: the role of cadherin adhesion and signaling in generating junctional contractility. *Curr. Top. Dev. Biol.* **112,** 65–102 (2015).

8. Huveneers, S. *et al.* Vinculin associates with endothelial VE-cadherin junctions to control force-dependent remodeling. *J. Cell Biol.* **196,** 641–52 (2012).

9. Wheelock, M. J., Shintani, Y., Maeda, M., Fukumoto, Y. & Johnson, K. R. Cadherin switching. *J. Cell Sci.* **121,** 727–35 (2008).

10. Sotomayor, M., Gaudet, R. & Corey, D. P. Sorting out a promiscuous superfamily: towards cadherin connectomics. *Trends Cell Biol.* **24,** 524–36 (2014).

11. van Roy, F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nat. Rev. Cancer* **14,** 121–34 (2014).

12. Berx, G. & van Roy, F. Involvement of Members of the Cadherin Superfamily in Cancer. *Cold Spring Harb. Perspect. Biol.* **1,** a003129–a003129 (2009).

13. van Roy, F. & Berx, G. The cell-cell adhesion molecule E-cadherin. *Cell. Mol. Life Sci.* **65,** 3756–88 (2008).

14. Gheldof, A. & Berx, G. Cadherins and epithelial-to-mesenchymal transition. *Prog. Mol. Biol. Transl. Sci.* **116,** 317–36 (2013).

15. Hirano, S. & Takeichi, M. Cadherins in brain morphogenesis and wiring. *Physiol. Rev.* **92,** 597–634 (2012).

16. Hirayama, T. & Yagi, T. Clustered protocadherins and neuronal diversity. *Prog Mol Biol Transl Sci* **116,** 145–167 (2013).

17. Mahoney, P. A. *et al.* The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily. *Cell* **67,** 853–868 (1991).

18. Sano, K. *et al.* Protocadherins: a large family of cadherin-related molecules in central nervous

system. *EMBO J* **12,** 2249–2256 (1993).

19. Oda, H., Uemura, T., Harada, Y., Iwai, Y. & Takeichi, M. A Drosophila homolog of cadherin associated with armadillo and essential for embryonic cell-cell adhesion. *Dev. Biol.* **165,** 716–26 (1994).

20. Noonan, J. P., Grimwood, J., Schmutz, J., Dickson, M. & Myers, R. M. Gene conversion and the evolution of protocadherin gene cluster diversity. *Genome Res.* **14,** 354–66 (2004).

21. Hulpiau, P. & van Roy, F. New insights into the evolution of metazoan cadherins. *Mol. Biol. Evol.* **28,** 647–57 (2011).

22. Vanhalst, K., Kools, P., Staes, K., van Roy, F. & Redies, C. delta-Protocadherins: a gene family expressed differentially in the mouse brain. *Cell Mol Life Sci* **62,** 1247–1259 (2005).

23. Takagi, J. Structural basis for ligand recognition by RGD (Arg-Gly-Asp)-dependent integrins. *Biochem. Soc. Trans.* **32,** 403–6 (2004).

24. Morishita, H. *et al.* Structure of the cadherin-related neuronal receptor/protocadherin-alpha first extracellular cadherin domain reveals diversity across cadherin families. *J. Biol. Chem.* **281,** 33650–63 (2006).

25. Yagi, T. Clustered protocadherin family. *Dev. Growth Differ.* **50 Suppl 1,** S131–40 (2008).

26. Esumi, S. *et al.* Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.* **37,** 171–6 (2005).

27. Tasic, B. *et al.* Promoter choice determines splice site selection in protocadherin alpha and gamma pre-mRNA splicing. *Mol Cell* **10,** 21–33 (2002).

28. Wang, X., Su, H. & Bradley, A. Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis-alternative splicing model. *Genes Dev* **16,** 1890–1905 (2002).

29. Morishita, H. & Yagi, T. Protocadherin family: diversity, structure, and function. *Curr. Opin. Cell Biol.* **19,** 584–92 (2007).

30. Kaneko, R. *et al.* Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J. Biol. Chem.* **281,** 30551–60 (2006).

31. Montague, P. R. & Friedlander, M. J. Morphogenesis and territorial coverage by isolated mammalian retinal ganglion cells. *J. Neurosci.* **11,** 1440–57 (1991).

32. Grueber, W. B. & Sagasti, A. Self-avoidance and tiling: Mechanisms of dendrite and axon spacing. *Cold Spring Harb. Perspect. Biol.* **2,** a001750 (2010).

33. Lefebvre, J. L., Kostadinov, D., Chen, W. V, Maniatis, T. & Sanes, J. R. Protocadherins mediate dendritic self-avoidance in the mammalian nervous system. *Nature* **488,** 517–521 (2012).

34. Zipursky, S. L. & Sanes, J. R. Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. *Cell* **143,** 343–353 (2010).

35. Chen, W. V & Maniatis, T. Clustered protocadherins. *Development* **140,** 3297–3302 (2013).

36. Zipursky, S. L. & Grueber, W. B. The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* **36,** 547–68 (2013).

37. Kostadinov, D. & Sanes, J. R. Protocadherin-dependent dendritic self-avoidance regulates neural connectivity and circuit function. *Elife* **4,** e08964 (2015).

38. Goodman, K. M. *et al.* Structural Basis of Diverse Homophilic Recognition by Clustered α- and β-Protocadherins. *Neuron* **90,** 709–723 (2016).

39. Schreiner, D. & Weiner, J. A. Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 14893–8 (2010).

40. Triana-Baltzer, G. B. & Blank, M. Cytoplasmic domain of protocadherin-alpha enhances homophilic interactions and recognizes cytoskeletal elements. *J. Neurobiol.* **66,** 393–407 (2006).

41. Hambsch, B., Grinevich, V., Seeburg, P. H. & Schwarz, M. K. {gamma}-Protocadherins, presenilin-mediated release of C-terminal fragment promotes locus expression. *J. Biol. Chem.* **280,** 15888–97 (2005).

42. Murata, Y., Hamada, S., Morishita, H., Mutoh, T. & Yagi, T. Interaction with protocadherin-gamma regulates the cell surface expression of protocadherin-alpha. *J. Biol. Chem.* **279,** 49508–16 (2004).

43. Han, M. H., Lin, C., Meng, S. & Wang, X. Proteomics analysis reveals overlapping functions of clustered protocadherins. *Mol Cell Proteomics* **9,** 71–83 (2010).

44. Thu, C. A. *et al.* Single-cell identity generated by combinatorial homophilic interactions between α, β, and γ protocadherins. *Cell* **158,** 1045–59 (2014).

45. Hayashi, S. & Takeichi, M. Emerging roles of protocadherins: from self-avoidance to enhancement of motility. *J. Cell Sci.* **128,** 1455–64 (2015).

46. Garrett, A. M., Schreiner, D., Lobas, M. A. & Weiner, J. A. gamma-protocadherins control cortical dendrite arborization by regulating the activity of a FAK/PKC/MARCKS signaling pathway. *Neuron* **74,** 269–276 (2012).

47. Rubinstein, R. *et al.* Molecular logic of neuronal self-recognition through protocadherin domain interactions. *Cell* **163,** 629–42 (2015).

48. Nicoludis, J. M. *et al.* Structure and Sequence Analyses of Clustered Protocadherins Reveal Antiparallel Interactions that Mediate Homophilic Specificity. *Structure* **23,** 2087–98 (2015).

49. Chen, J. *et al.* alpha- and gamma-Protocadherins negatively regulate PYK2. *J Biol Chem* **284,** 2880–2890 (2009).

50. Emond, M. R. & Jontes, J. D. Inhibition of protocadherin-alpha function results in neuronal death in the developing zebrafish. *Dev. Biol.* **321,** 175–87 (2008).

51. Keeler, A. B., Schreiner, D. & Weiner, J. A. Protein Kinase C Phosphorylation of a γ-Protocadherin C-terminal Lipid Binding Domain Regulates Focal Adhesion Kinase Inhibition and Dendrite Arborization. *J. Biol. Chem.* **290,** 20674–86 (2015).

52. Williams, N. A., Close, J. P., Giouzeli, M. & Crow, T. J. Accelerated evolution of Protocadherin11X/Y: a candidate gene-pair for cerebral asymmetry and language. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **141B,** 623–33 (2006).

53. Blanco-Arias, P., Sargent, C. A. & Affara, N. A. Protocadherin X (PCDHX) and Y (PCDHY) genes; multiple mRNA isoforms encoding variant signal peptides and cytoplasmic domains. *Mamm. Genome* **15,** 41–52 (2004).

54. Blanco, P., Sargent, C. A., Boucher, C. A., Mitchell, M. & Affara, N. A. Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain. *Mamm. Genome* **11,** 906–914 (2000).

55. Koppelman, G. H. *et al.* Identification of PCDH1 as a novel susceptibility gene for bronchial hyperresponsiveness. *Am. J. Respir. Crit. Care Med.* **180,** 929–35 (2009).

56. Fukata, Y. & Fukata, M. Protein palmitoylation in neuronal development and synaptic plasticity. *Nat. Rev. Neurosci.* **11,** 161–75 (2010).

57. Sano, K. *et al.* Protocadherins: a large family of cadherin-related molecules in central nervous system. *EMBO J.* **12,** 2249–56 (1993).

58. Koning, H. *et al.* Characterization of protocadherin-1 expression in primary bronchial epithelial cells: association with epithelial cell differentiation. *FASEB J.* **26,** 439–48 (2012).

59. Uemura, M., Nakao, S., Suzuki, S. T., Takeichi, M. & Hirano, S. OL-Protocadherin is essential for growth of striatal axons and thalamocortical projections. *Nat. Neurosci.* **10,** 1151–9 (2007).

60. Wolverton, T. & Lalande, M. Identification and characterization of three members of a novel subclass of protocadherins. *Genomics* **76,** 66–72 (2001).

61. Ying, J. *et al.* Functional epigenetics identifies a protocadherin PCDH10 as a candidate tumor suppressor for nasopharyngeal, esophageal and multiple other carcinomas with frequent methylation. *Oncogene* **25,** 1070–80 (2006).

62. Kahr, I., Vandepoele, K. & van Roy, F. Delta-protocadherins in health and disease. *Prog Mol Biol Transl Sci* **116,** 169–192 (2013).

63. Lin, J., Wang, C. & Redies, C. Expression of delta-protocadherins in the spinal cord of the chicken embryo. *J Comp Neurol* **520,** 1509–1531 (2012).

64. Lin, J. *et al.* Anatomical expression patterns of delta-protocadherins in developing chicken cochlea. *J Anat* **221,** 598–608 (2012).

65. Hertel, N. & Redies, C. Absence of layer-specific cadherin expression profiles in the neocortex of the reeler mutant mouse. *Cereb. Cortex* **21,** 1105–17 (2011).

66. Kim, S.-Y., Chung, H. S., Sun, W. & Kim, H. Spatiotemporal expression pattern of non-clustered protocadherin family members in the developing rat brain. *Neuroscience* **147,** 996–1021 (2007).

67. Krishna-K, K., Hertel, N. & Redies, C. Cadherin expression in the somatosensory cortex: evidence for a combinatorial molecular code at the single-cell level. *Neuroscience* **175,** 37–48

(2011).

68. Kim, S. Y. *et al.* The expression of non-clustered protocadherins in adult rat hippocampal formation and the connecting brain regions. *Neuroscience* **170,** 189–99 (2010).

69. Aamar, E. & Dawid, I. B. Protocadherin-18a has a role in cell adhesion, behavior and migration in zebrafish development. *Dev. Biol.* **318,** 335–346 (2008).

70. Kuroda, H., Inui, M., Sugimoto, K., Hayata, T. & Asashima, M. Axial protocadherin is a mediator of prenotochord cell sorting in Xenopus. *Dev. Biol.* **244,** 267–77 (2002).

71. Yamagata, K. *et al.* Arcadlin is a neural activity-regulated cadherin involved in long term potentiation. *J. Biol. Chem.* **274,** 19473–1979 (1999).

72. Hirano, S., Yan, Q. & Suzuki, S. T. Expression of a novel protocadherin, OL-protocadherin, in a subset of functional systems of the developing mouse brain. *J. Neurosci.* **19,** 995–1005 (1999).

73. Yoshida, K., Yoshitomo-Nakagawa, K., Seki, N., Sasaki, M. & Sugano, S. Cloning, expression analysis, and chromosomal localization of BH-protocadherin (PCDH7), a novel member of the cadherin superfamily. *Genomics* **49,** 458–461 (1998).

74. Yoshida, K. Fibroblast cell shape and adhesion in vitro is altered by overexpression of the 7a and 7b isoforms of protocadherin 7, but not the 7c isoform. *Cell. Mol. Biol. Lett.* **8,** 735–41 (2003).

75. Bradley, R. S., Espeseth, A. & Kintner, C. NF-protocadherin, a novel member of the cadherin superfamily, is required for Xenopus ectodermal differentiation. *Curr. Biol.* **8,** 325–34 (1998).

76. Kim, S. H., Yamamoto, A., Bouwmeester, T., Agius, E. & Robertis, E. M. The role of paraxial protocadherin in selective adhesion and cell movements of the mesoderm during Xenopus gastrulation. *Development* **125,** 4681–4690 (1998).

77. Tai, K., Kubota, M., Shiono, K., Tokutsu, H. & Suzuki, S. T. Adhesion properties and retinofugal expression of chicken protocadherin-19. *Brain Res.* **1344,** 13–24 (2010).

78. Chen, X. & Gumbiner, B. M. Paraxial protocadherin mediates cell sorting and tissue morphogenesis by regulating C-cadherin adhesion activity. *J. Cell Biol.* **174,** 301–313 (2006).

79. Chen, X., Koh, E., Yoder, M. & Gumbiner, B. M. A protocadherin-cadherin-FLRT3 complex controls cell adhesion and morphogenesis. *PLoS One* **4,** e8411 (2009).

80. Yasuda, S. *et al.* Activity-induced protocadherin arcadlin regulates dendritic spine number by triggering N-cadherin endocytosis via TAO2beta and p38 MAP kinases. *Neuron* **56,** 456–71 (2007).

81. Biswas, S., Emond, M. R. & Jontes, J. D. Protocadherin-19 and N-cadherin interact to control cell movements during anterior neurulation. *J. Cell Biol.* **191,** 1029–41 (2010).

82. Emond, M. R., Biswas, S., Blevins, C. J. & Jontes, J. D. A complex of Protocadherin-19 and N-cadherin mediates a novel mechanism of cell adhesion. *J. Cell Biol.* **195,** 1115–21 (2011).

83. Emond, M. R., Biswas, S. & Jontes, J. D. Protocadherin-19 is essential for early steps in brain morphogenesis. *Dev. Biol.* **334,** 72–83 (2009).

84. Blevins, C. J., Emond, M. R., Biswas, S. & Jontes, J. D. Differential expression, alternative splicing, and adhesive properties of the zebrafish delta1-protocadherins. *Neuroscience* **199,** 523–534 (2011).

85. Nakao, S., Platek, A., Hirano, S. & Takeichi, M. Contact-dependent promotion of cell migration by the OL-protocadherin-Nap1 interaction. *J. Cell Biol.* **182,** 395–410 (2008).

86. Demontis, F., Habermann, B. & Dahmann, C. PDZ-domain-binding sites are common among cadherins. *Dev. Genes Evol.* **216,** 737–41 (2006).

87. Kai, M. *et al.* Phosphorylation-Dependent Ubiquitination of Paraxial Protocadherin (PAPC) Controls Gastrulation Cell Movements. *PLoS One* **10,** e0115111 (2015).

88. Yoshida, K., Watanabe, M., Kato, H., Dutta, A. & Sugano, S. BH-protocadherin-c, a member of the cadherin superfamily, interacts with protein phosphatase 1 alpha through its intracellular domain. *FEBS Lett.* **460,** 93–98 (1999).

89. Jérôme, M. & Paudel, H. K. 14-3-3ζ regulates nuclear trafficking of protein phosphatase 1α (PP1α) in HEK-293 cells. *Arch. Biochem. Biophys.* **558,** 28–35 (2014).

90. da Cruz e Silva, E. F. *et al.* Differential expression of protein phosphatase 1 isoforms in mammalian brain. *J. Neurosci.* **15,** 3375–89 (1995).

91. Ouimet, C. C., da Cruz e Silva, E. F. & Greengard, P. The alpha and gamma 1 isoforms of protein phosphatase 1 are highly and specifically concentrated in dendritic spines. *Proc. Natl. Acad. Sci.* **92,** 3396–3400 (1995).

92. Hemmings, H. C., Greengard, P., Tung, H. Y. L. & Cohen, P. DARPP-32, a dopamine-regulated neuronal phosphoprotein, is a potent inhibitor of protein phosphatase-1. *Nature* **310,** 503–505 (1984).

93. Halford, M. M. & Stacker, S. A. Revelations of the RYK receptor. *Bioessays* **23,** 34–45 (2001).

94. Lin, S., Baye, L. M., Westfall, T. A. & Slusarski, D. C. Wnt5b-Ryk pathway provides directional signals to regulate gastrulation movement. *J. Cell Biol.* **190,** 263–78 (2010).

95. Kim, G.-H., Her, J.-H. & Han, J.-K. Ryk cooperates with Frizzled 7 to promote Wnt11-mediated endocytosis and is essential for Xenopus laevis convergent extension movements. *J. Cell Biol.* **182,** 1073–82 (2008).

96. Medina, A., Swain, R. K., Kuerner, K.-M. & Steinbeisser, H. Xenopus paraxial protocadherin has signaling functions and is involved in tissue separation. *EMBO J.* **23,** 3249–58 (2004).

97. Schambony, A. & Wedlich, D. Wnt-5A/Ror2 regulate expression of XPAPC through an alternative noncanonical signaling pathway. *Dev. Cell* **12,** 779–92 (2007).

98. Colland, F. *et al.* Functional proteomics mapping of a human signaling pathway. *Genome Res.* **14,** 1324–32 (2004).

99. Faura Tellez, G. *et al.* Protocadherin-1 binds to SMAD3 and suppresses TGF-β1-induced gene transcription. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **309,** L725–35 (2015).

100. Heggem, M. A. & Bradley, R. S. The cytoplasmic domain of Xenopus NF-protocadherin interacts with TAF1/set. *Dev. Cell* **4,** 419–29 (2003).

101. Rashid, D., Newell, K., Shama, L. & Bradley, R. A requirement for NF-protocadherin and TAF1/Set in cell adhesion and neural tube formation. *Dev. Biol.* **291,** 170–81 (2006).

102. Piper, M., Dwivedy, A., Leung, L., Bradley, R. S. & Holt, C. E. NF-protocadherin and TAF1 regulate retinal axon initiation and elongation in vivo. *J. Neurosci.* **28,** 100–5 (2008).

103. Leung, L. C., Harris, W. A., Holt, C. E. & Piper, M. NF-Protocadherin Regulates Retinal Ganglion Cell Axon Behaviour in the Developing Visual System. *PLoS One* **10,** e0141290 (2015).

104. Pollitt, A. Y. & Insall, R. H. WASP and SCAR/WAVE proteins: the drivers of actin assembly. *J. Cell Sci.* **122,** 2575–8 (2009).

105. Takenawa, T. & Suetsugu, S. The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. *Nat. Rev. Mol. Cell Biol.* **8,** 37–48 (2007).

106. Yokota, Y., Ring, C., Cheung, R., Pevny, L. & Anton, E. S. Nap1-regulated neuronal cytoskeletal dynamics is essential for the final differentiation of neurons in cerebral cortex. *Neuron* **54,** 429–45 (2007).

107. Chen, B. *et al.* The WAVE regulatory complex links diverse receptors to the actin cytoskeleton. *Cell* **156,** 195–207 (2014).

108. Chen, J., Yu, S., Fu, Y. & Li, X. Synaptic proteins and receptors defects in autism spectrum disorders. *Front. Cell. Neurosci.* **8,** 276 (2014).

109. Biswas, S. *et al.* Protocadherin-18b interacts with Nap1 to control motor axon growth and arborization in zebrafish. *Mol. Biol. Cell* **25,** 633–42 (2014).

110. Howell, B. W., Gertler, F. B. & Cooper, J. A. Mouse disabled (mDab1): a Src binding protein implicated in neuronal development. *EMBO J.* **16,** 121–32 (1997).

111. Sheldon, M. *et al.* Scrambler and yotari disrupt the disabled gene and produce a reeler-like phenotype in mice. *Nature* **389,** 730–3 (1997).

112. D'Arcangelo, G. *et al.* A protein related to extracellular matrix proteins deleted in the mouse mutant reeler. *Nature* **374,** 719–23 (1995).

113. Rice, D. S. *et al.* Disabled-1 acts downstream of Reelin in a signaling pathway that controls laminar organization in the mammalian brain. *Development* **125,** 3719–29 (1998).

114. Homayouni, R., Rice, D. S. & Curran, T. Disabled-1 interacts with a novel developmentally regulated protocadherin. *Biochem. Biophys. Res. Commun.* **289,** 539–47 (2001).

115. Wang, Y. *et al.* Xenopus Paraxial Protocadherin regulates morphogenesis by antagonizing Sprouty. *Genes Dev.* **22,** 878–83 (2008).

116. Chung, H. A., Yamamoto, T. S. & Ueno, N. ANR5, an FGF target gene product, regulates gastrulation in Xenopus. *Curr. Biol.* **17,** 932–9 (2007).

117. Kietzmann, A., Wang, Y., Weber, D. & Steinbeisser, H. Xenopus paraxial protocadherin inhibits Wnt/β-catenin signalling via casein kinase 2β. *EMBO Rep.* **13,** 129–34 (2012).

118. Unterseher, F. *et al.* Paraxial protocadherin coordinates cell polarity during convergent extension via Rho A and JNK. *EMBO J.* **23,** 3259–69 (2004).

119. Martinez, S. *et al.* The PTK7 and ROR2 receptors interact in the vertebrate WNT/PCP pathway. *J. Biol. Chem.* (2015). doi:10.1074/jbc.M115.697615

120. Yamamoto, A., Kemp, C., Bachiller, D., Geissert, D. & De Robertis, E. M. Mouse paraxial protocadherin is expressed in trunk mesoderm and is not essential for mouse development. *Genesis* **27,** 49–57 (2000).

121. Rhee, J., Takahashi, Y., Saga, Y., Wilson-Rawls, J. & Rawls, A. The protocadherin papc is involved in the organization of the epithelium along the segmental border during mouse somitogenesis. *Dev. Biol.* **254,** 248–61 (2003).

122. Compagnucci, C. *et al.* Characterizing PCDH19 in human induced pluripotent stem cells (iPSCs) and iPSC-derived developing neurons: emerging role of a protein involved in controlling polarity during neurogenesis. *Oncotarget* **6,** 26804–26813 (2015).

123. Miyake, K. *et al.* The protocadherins, PCDHB1 and PCDH7, are regulated by MeCP2 in neuronal cells and brain tissues: implication for pathogenesis of Rett syndrome. *BMC Neurosci* **12,** 81 (2011).

124. Williams, E. O. *et al.* Delta Protocadherin 10 is Regulated by Activity in the Mouse Main Olfactory System. *Front. Neural Circuits* **5,** 9 (2011).

125. Caselli, R. *et al.* Retinoblastoma and mental retardation microdeletion syndrome: clinical characterization and molecular dissection using array CGH. *J. Hum. Genet.* **52,** 535–542 (2007).

126. Speevak, M. D. & Farrell, S. A. Non-syndromic language delay in a child with disruption in the Protocadherin11X/Y gene pair. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **156B,** 484–9 (2011).

127. Carrasquillo, M. M. *et al.* Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat. Genet.* **41,** 192–8 (2009).

128. Kasnauskiene, J. *et al.* A single gene deletion on 4q28.3: PCDH18 – A new candidate gene for intellectual disability? *Eur. J. Med. Genet.* **55,** 274–277 (2012).

129. Dibbens, L. M. *et al.* X-linked protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment. *Nat. Genet.* **40,** 776–81 (2008).

130. Dror, N. *et al.* Identification of IRF-8 and IRF-1 target genes in activated macrophages. *Mol. Immunol.* **44,** 338–46 (2007).

131. Li, X. *et al.* IKKalpha, IKKbeta, and NEMO/IKKgamma are each required for the NF-kappa B-mediated inflammatory response program. *J. Biol. Chem.* **277,** 45129–40 (2002).

132. Vazquez-Cintron, E. J. *et al.* Protocadherin-18 is a novel differentiation marker and an inhibitory signaling receptor for CD8+ effector memory T cells. *PLoS One* **7,** e36101 (2012).

133. Koneru, M., Schaer, D., Monu, N., Ayala, A. & Frey, A. B. Defective proximal TCR signaling inhibits CD8+ tumor-infiltrating lymphocyte lytic function. *J. Immunol.* **174,** 1830–40 (2005).

134. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–6 (2008).

135. Miyamoto, K. *et al.* Identification of 20 genes aberrantly methylated in human breast cancers. *Int. J. Cancer* **116,** 407–14 (2005).

136. Ying, J. *et al.* Frequent epigenetic silencing of protocadherin 10 by methylation in multiple haematologic malignancies. *Br. J. Haematol.* **136,** 829–32 (2007).

137. Imoto, I. *et al.* Frequent silencing of the candidate tumor suppressor PCDH20 by epigenetic mechanism in non-small-cell lung cancers. *Cancer Res.* **66,** 4617–26 (2006).

138. Haruki, S. *et al.* Frequent silencing of protocadherin 17, a candidate tumour suppressor for esophageal squamous cell carcinoma. *Carcinogenesis* **31,** 1027–36 (2010).

139. Yu, J. *et al.* Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* **136,** 640–51.e1 (2009).

140. Wang, C. *et al.* Downregulation of PCDH9 predicts prognosis for patients with glioma. *J. Clin. Neurosci.* **19,** 541–5 (2012).

141. Narayan, G. *et al.* Promoter methylation-mediated inactivation of PCDH10 in acute lymphoblastic leukemia contributes to chemotherapy resistance. *Genes, Chromosom. Cancer* **50,** 1043–1053 (2011).

142. Neben, K. *et al.* Microarray-based screening for molecular markers in medulloblastoma revealed STK15 as independent predictor for survival. *Cancer Res.* **64,** 3103–11 (2004).

143. Vincent, A. *et al.* Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. *Clin. Cancer Res.* **17,** 4341–54 (2011).

144. Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459,** 1005–9 (2009).

145. Li, A.-M. *et al.* Protocadherin-7 induces bone metastasis of breast cancer. *Biochem. Biophys. Res. Commun.* **436,** 486–90 (2013).

146. Suzuki, H. *et al.* A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat. Genet.* **31,** 141–9 (2002).

147. Leshchenko, V. V *et al.* Genomewide DNA methylation analysis reveals novel targets for drug development in mantle cell lymphoma. *Blood* **116,** 1025–34 (2010).

148. Morris, M. R. *et al.* Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. *Oncogene* **30,** 1390–401 (2011).

149. He, D. *et al.* Protocadherin8 is a functional tumor suppressor frequently inactivated by promoter methylation in nasopharyngeal carcinoma. *Eur. J. Cancer Prev.* **21,** 569–75 (2012).

150. Yu, J. S. *et al.* PCDH8, the human homolog of PAPC, is a candidate tumor suppressor of breast cancer. *Oncogene* **27,** 4657–65 (2008).

151. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–4 (2012).

152. de Tayrac, M. *et al.* Integrative genome-wide analysis reveals a robust genomic glioblastoma signature associated with copy number driving changes in gene expression. *Genes.*

Chromosomes Cancer **48,** 55–68 (2009).

153. Wang, C. *et al.* Characterizing the Role of PCDH9 in the Regulation of Glioma Cell Apoptosis and Invasion. *J Mol Neurosci* (2013). doi:10.1007/s12031-013-0133-2

154. Bamford, S. *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* **91,** 355–8 (2004).

155. Narayan, G. *et al.* Protocadherin PCDH10, involved in tumor progression, is a frequent and early target of promoter hypermethylation in cervical cancer. *Genes. Chromosomes Cancer* **48,** 983–92 (2009).

156. Cheung, H. H. *et al.* Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. *Br. J. Cancer* **102,** 419–27 (2010).

157. Li, Y., Yang, Z., Song, J., Liu, Q. & Chen, J. Protocadherin-10 is involved in angiogenesis and methylation correlated with multiple myeloma. *Int. J. Mol. Med.* **29,** 704–10 (2012).

158. Li, Z. *et al.* Epigenetic inactivation of PCDH10 in human prostate cancer cell lines. *Cell Biol. Int.* **35,** 671–6 (2011).

159. Li, Z. *et al.* Role of PCDH10 and its hypermethylation in human gastric cancer. *Biochim. Biophys. Acta* **1823,** 298–305 (2012).

160. Bertrand, K. C. *et al.* PCDH10 is a candidate tumour suppressor gene in medulloblastoma. *Childs. Nerv. Syst.* **27,** 1243–9 (2011).

161. Kumar, A. *et al.* Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 17087–92 (2011).

162. Chen, M.-W. *et al.* The emergence of protocadherin-PC expression during the acquisition of apoptosis-resistance by prostate cancer cells. *Oncogene* **21,** 7861–71 (2002).

163. Yang, Y., Liu, J., Li, X. & Li, J.-C. PCDH17 gene promoter demethylation and cell cycle arrest by genistein in gastric cancer. *Histol. Histopathol.* **27,** 217–24 (2012).

164. Hu, X. *et al.* Protocadherin 17 acts as a tumour suppressor inducing tumour cell apoptosis and autophagy, and is frequently methylated in gastric and colorectal cancers. *J. Pathol.* **229,** 62–73 (2013).

165. American Cancer Society. *Cancer Facts & Figures 2015*. (2015).

166. Morris, M. R. & Maher, E. R. Epigenetics of renal cell carcinoma: the path towards new diagnostics and therapeutics. *Genome Med.* **2,** 59 (2010).

167. Fang, S. *et al.* Silencing of PCDH10 in hepatocellular carcinoma via de novo DNA methylation independent of HBV infection or HBX expression. *Clin. Exp. Med.* **13,** 127–34 (2013).

168. Zhao, Y. *et al.* A novel wnt regulatory axis in endometrioid endometrial cancer. *Cancer Res.* **74,** 5103–17 (2014).

169. Xu, Y. *et al.* PCDH10 inhibits cell proliferation of multiple myeloma via the negative regulation of the Wnt/β-catenin/BCL-9 signaling pathway. *Oncol. Rep.* **34,** 747–754 (2015).

170. Li, Z. *et al.* Nuclear factor-κB is involved in the protocadherin-10-mediated pro-apoptotic effect in multiple myeloma. *Mol. Med. Rep.* **10,** 832–8 (2014).

171. Jao, T.-M. *et al.* Protocadherin 10 suppresses tumorigenesis and metastasis in colorectal cancer and its genetic loss predicts adverse prognosis. *Int. J. Cancer* **135,** 2593–603 (2014).

172. Ding, K. *et al.* Inhibition of apoptosis by downregulation of hBex1, a novel mechanism, contributes to the chemoresistance of Bcr/Abl+ leukemic cells. *Carcinogenesis* **30,** 35–42 (2009).

173. Wang, K.-H., Liu, H.-W., Lin, S.-R., Ding, D.-C. & Chu, T.-Y. Field methylation silencing of the protocadherin 10 gene in cervical carcinogenesis as a potential specific diagnostic test from cervical scrapings. *Cancer Sci.* **100,** 2175–80 (2009).

174. Giefing, M. *et al.* High resolution ArrayCGH and expression profiling identifies PTPRD and PCDH17/PCH68 as tumor suppressor gene candidates in laryngeal squamous cell carcinoma. *Genes. Chromosomes Cancer* **50,** 154–66 (2011).

175. Wu, J.-C. *et al.* Se-Allylselenocysteine induces autophagy by modulating the AMPK/mTOR signaling pathway and epigenetic regulation of PCDH17 in human colorectal adenocarcinoma cells. *Mol. Nutr. Food Res.* (2015). doi:10.1002/mnfr.201500373

176. Dang, Z. *et al.* Loss of protocadherin-17 (PCDH-17) promotes metastasis and invasion through hyperactivation of EGFR/MEK/ERK signaling pathway in hepatocellular carcinoma. *Tumour Biol.* (2015). doi:10.1007/s13277-015-3970-5

177. Hernan, R. *et al.* ERBB2 up-regulates S100A4 and several other prometastatic genes in medulloblastoma. *Cancer Res.* **63,** 140–8 (2003).

178. Singh, K., Kang, P. J. & Park, H.-O. The Rho5 GTPase is necessary for oxidant-induced cell death in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 1522–7 (2008).

179. Jaeger, J. *et al.* Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. *Clin. Cancer Res.* **13,** 806–15 (2007).

180. Mittal, A. K. *et al.* Molecular basis of aggressive disease in chronic lymphocytic leukemia patients with 11q deletion and trisomy 12 chromosomal abnormalities. *Int. J. Mol. Med.* **20,** 461–9 (2007).

181. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* **9,** 341–55 (2008).

182. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (2013).

183. Hallmayer, J. *et al.* Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* **68,** 1095–102 (2011).

184. Betancur, C., Sakurai, T. & Buxbaum, J. D. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. *Trends Neurosci.* **32,** 402–12 (2009).

185. Toro, R. *et al.* Key role for gene dosage and synaptic homeostasis in autism spectrum

disorders. *Trends Genet.* **26,** 363–72 (2010).

186. Lai, M.-C., Lombardo, M. V & Baron-Cohen, S. Autism. *Lancet* **383,** 896–910 (2014).

187. Marshall, C. R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 477–88 (2008).

188. Bucan, M. *et al.* Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. *PLoS Genet.* **5,** e1000536 (2009).

189. Girirajan, S. *et al.* Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *Am. J. Hum. Genet.* **92,** 221–37 (2013).

190. Leblond, C. S. *et al.* Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. *PLoS Genet.* **8,** e1002521 (2012).

191. Luo, R. *et al.* Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *Am. J. Hum. Genet.* **91,** 38–55 (2012).

192. Bruining, H. *et al.* Genetic Mapping in Mice Reveals the Involvement of Pcdh9 in Long-Term Social and Object Recognition and Sensorimotor Development. *Biol. Psychiatry* **78,** 485–95 (2015).

193. Morrow, E. M. *et al.* Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321,** 218–23 (2008).

194. Dastjerdi, F. V, Consalez, G. G. & Hawkes, R. Pattern formation during development of the embryonic cerebellum. *Front. Neuroanat.* **6,** 10 (2012).

195. Redies, C., Neudert, F. & Lin, J. Cadherins in cerebellar development: translation of embryonic patterning into mature functional compartmentalization. *Cerebellum* **10,** 393–408 (2011).

196. Frank, M. & Kemler, R. Protocadherins. *Curr Opin Cell Biol* **14,** 557–562 (2002).

197. Camacho, A. *et al.* Cognitive and behavioral profile in females with epilepsy with PDCH19 mutation: two novel mutations and review of the literature. *Epilepsy Behav.* **24,** 134–7 (2012).

198. Piton, A. *et al.* Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. *Mol. Psychiatry* **16,** 867–80 (2011).

199. Pham, D., Tan, C., Homan, C., Jolly, L. & Gecz, J. in *Neuronal Synaptic Dysfunct. Autism Spectr. Disord. Intellect. Disabil.* 221–231 (2016). doi:10.1016/B978-0-12-800109-7.00014-5

200. Rett, A. [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. *Wien. Med. Wochenschr.* **116,** 723–6 (1966).

201. Hagberg, B., Aicardi, J., Dias, K. & Ramos, O. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann. Neurol.* **14,** 471–9 (1983).

202. Chen, R. Z., Akbarian, S., Tudor, M. & Jaenisch, R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* **27,** 327–31 (2001).

203. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27,** 322–6 (2001).

204. Webster, J. P., Kaushik, M., Bristow, G. C. & McConkey, G. A. Toxoplasma gondii infection, from predation to schizophrenia: can animal behaviour help us understand human behaviour? *J. Exp. Biol.* **216,** 99–112 (2013).

205. Kallmann, F. J. The genetics of schizophrenia. (1938).

206. van Os, J. & Kapur, S. Schizophrenia. *Lancet* **374,** 635–45 (2009).

207. Strehl, S., Glatt, K., Liu, Q. M., Glatt, H. & Lalande, M. Characterization of two novel protocadherins (PCDH8 and PCDH9) localized on human chromosome 13 and mouse chromosome 14. *Genomics* **53,** 81–89 (1998).

208. Shaw, S. H. *et al.* A genome-wide search for schizophrenia susceptibility genes. *Am. J. Med. Genet.* **81,** 364–76 (1998).

209. Bray, N. J. *et al.* Screening the human protocadherin 8 (PCDH8) gene in schizophrenia. *Genes. Brain. Behav.* **1,** 187–91 (2002).

210. Dean, B., Keriakous, D., Scarr, E. & Thomas, E. A. Gene expression profiling in Brodmann's area 46 from subjects with schizophrenia. *Aust. N. Z. J. Psychiatry* **41,** 308–20 (2007).

211. Durand, C. M. *et al.* Expression and genetic variability ofPCDH11Y, a gene specific toHomo sapiens and candidate for susceptibility to psychiatric disorders. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **141B,** 67–70 (2006).

212. Scheffer, I. E. *et al.* Epilepsy and mental retardation limited to females: an under-recognized disorder. *Brain* **131,** 918–27 (2008).

213. Marini, C. *et al.* Protocadherin 19 mutations in girls with infantile-onset epilepsy. *Neurology* **75,** 646–653 (2010).

214. Depienne, C. & LeGuern, E. *PCDH19* -related infantile epileptic encephalopathy: An unusual X-linked inheritance disorder. *Hum. Mutat.* **33,** 627–634 (2012).

215. Koch, A. W. *et al.* Structure of the neural (N-) cadherin prodomain reveals a cadherin extracellular domain-like fold without adhesive characteristics. *Structure* **12,** 793–805 (2004).

Chapter II. How to investigate Protein-Protein Interactions

II.1 An introduction to Systems Biology and Interactomics

In the common view, biological systems are complex but consisting of simple and similar elements that cooperate to yield a complex behavior. However, this is not a realistic view: biological systems are actually formed by a large amount of different kinds of elements with different characteristics and functions that interact with each other in a selective way. Biological systems have functions that rely on combinations and mutual interactions of the different parts involved. Consequently, in living systems, for molecules, cells, organisms or even entire species, any prediction of their function is impossible to derive from the properties of the individual parts only. This forms the basis for the so-called System Biology approach.

Sir Paul Nurse identified four of what he calls 'the great ideas of biology': (i) the gene is the basis for heredity, (ii) the cell is the fundamental unit of organisms, (iii) biology is based on chemistry, and (iv) species evolve by natural selection[1]. This is how biology has been taught, how studies have been designed and carried out and this is the base for therapeutic approaches and for the way of thinking in modern science. Different study fields have been developed from these "great ideas" such as genetics and molecular biology, cell biology, biochemistry and evolutionary biology, respectively, giving four different approaches and points of view. Systems Biology is the "fifth great idea". Even though the Systems Biology discipline is today still developing, its theoretical basis dates back to more than a half century ago[2]. Already at the end of the 18th century, Immanuel Kant wrote about Systems Biology without giving it a name: *Critique of Judgment,* written in 1790, might be defined as the first philosophical treatise on Systems Biology. In this book, Kant perfectly described the principle of organization and collaboration of systems: "Organisms are organized natural products in which every part is reciprocally both end and means"[3].

Despite the fact that the theoretical basis of Systems Biology is relatively old, its application to a real field of study needed some time. In 1999, Leland H. Hartwell and colleagues published the paper "From molecular to modular cell biology", in Nature. There, they propose to combine molecular analysis / procedure and concepts from informatics: this was the onset of Systems Biology[4]. Since then, the development of new molecular technologies contributed significantly to the development of Systems Biology (**Figure 12**).

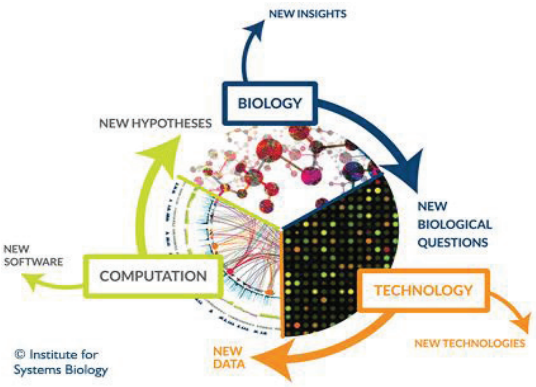


Figure 12 Systems Biology: Biology drives technology drives computation.
Figure from https://www.systemsbiology.org/about/what-is-systems-biology/

The concept of the cell changed dramatically: the focus is dedicated to the overall biological functions, which in a biological organization can be found at different levels, from organelles to organisms. In this view the cell acquires an important role as the simplest unit which still exhibits the characteristics of life[4]. A cell consists of hundreds of organized modules (e.g., the proteasome or the RNA exosome), composed of both macromolecules and small molecules. The interaction between those modules is termed "molecular sociology"[5]. The complex web of all the macromolecular interactions within and between cells constitutes the "interactome" network[6]. But thinking about Systems Biology as a single discipline would be a mistake: its power lies in its interdisciplinary concept. It brings together biologists, chemists, mathematicians, physicists, statisticians, computer scientists and engineers. Their respective specialties allow research to go through the "Omics Cascade": Systems Biology can be considered as the study of how the genome, the transcriptome, the proteome and the metabolome are regulated to determine a defined phenotype (**Figure 13**).

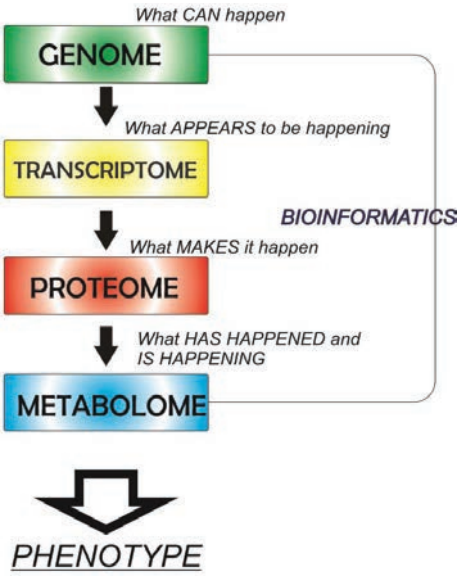


Figure 13 Systems Biology: From the Genome to the Phenotype

Over the years, the concept of "protein function" changed dramatically (**Figure 14**).

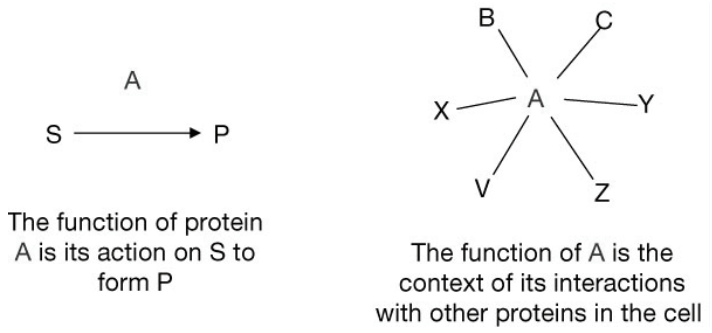


Figure 14 Protein function definition
On the left, the traditional view of protein function. On the right, the post-genomic consideration of protein function. Figure taken from[7].

The traditional view focuses on the action of a single protein molecule: the most representative example is the catalysis of a reaction. Within the Systems Biology context, this evolved towards what is called "cellular function" or "contextual function", to underline the role of each protein in a network of interacting molecules. Today, what defines the post-genomic era is the need of an expanded view on protein functions in a cellular and functional context. The development of computational methods to understand functional connections between proteins has been greatly facilitated by the advent of fully sequenced genomes[7] (Representative examples are shown in **Figure 15**).

Every protein can interact with metabolites, with other proteins and/or with nucleic acids. While the interaction with metabolites provides the cells with essential metabolic fluxes, the protein-protein interactions (PPIs) and the protein-nucleic acid interactions contribute to organizing, expressing and regulating cellular functions. Despite the fact that proteins can act as vital macromolecules at the cellular and systemic level, they rarely act alone, and if two proteins interact with each other, they usually play a role in the same cellular functions. This is why PPI studies are frequently used as a starting point for investigation of unknown proteins according to the principle of 'guilt by association': showing that a protein with unknown function interacts with one with known function gives a hint on the possible role of this unknown protein[8]. It has been estimated that there are about 400,000 PPI in the human interactome[9].

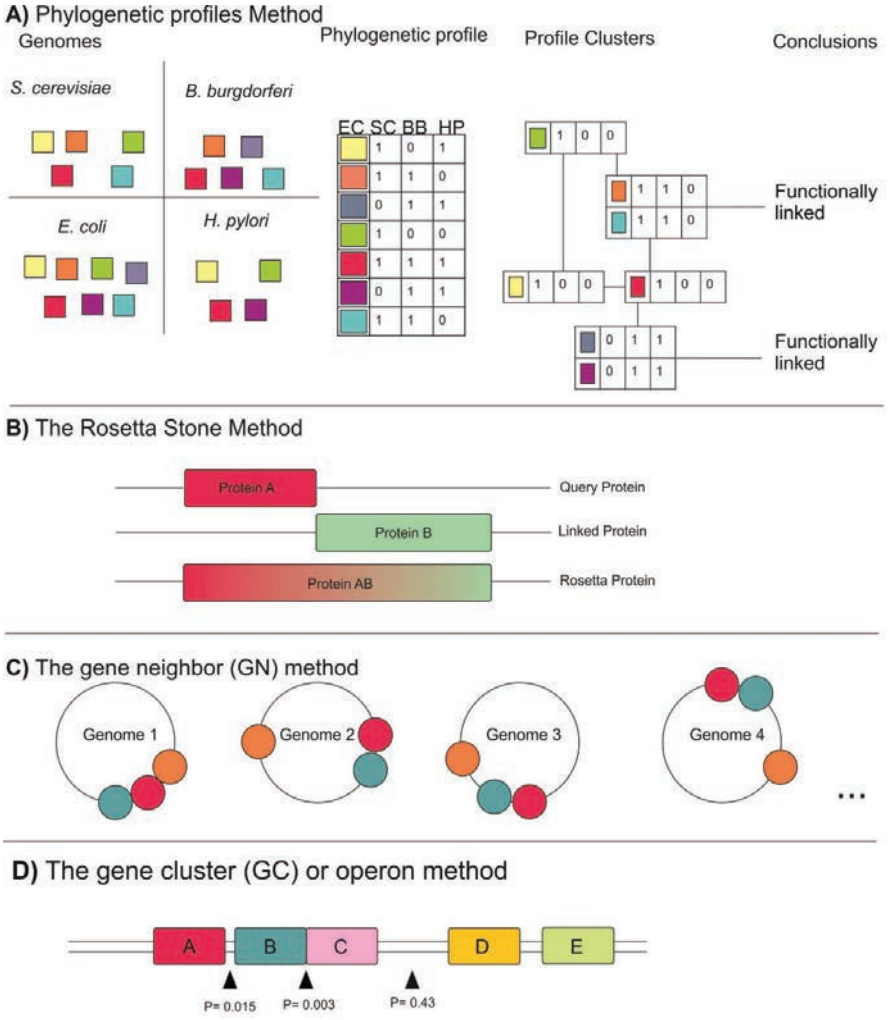


Figure 15 The general mechanism of four representative methods

A) Phylogenetic profile method: 4 different genomes with a set of proteins (colored squares), in the phylogenetic profile are represented with 1 if present, 0 if absent. Identical profiles are clustered in boxes on the right, with profiles differing by one bit connected by lines. Orange and light blue, and blue and purple are functionally linked because they have the same phylogenetic profile. B) The Rosetta stone concept or the domain fusion model searches for gene fusion events: protein A and protein B belong to the same organism. Protein AB is from another organism; the fused domain or Rosetta stone sequence it is homologous to two separate sequences in another species. For example the fused sequence of TrpC (protein AB) in the Escherichia coli genome would inform us that the yeast proteins TrG (protein A) and TrpF (protein B) are functionally linked. C) The gene neighbor method assume that If two genes (blue and red in the figure) are found to be neighbors in several different genomes, a functional linkage may be inferred between the proteins they encode. D) The gene cluster (GC) or operon method identifies closely spaced genes, and assigns a probability P of observing a particular gap distance (or smaller), as judged by the collective set of inter-gene distances[10].

II.2 Protein-Protein Interactions (PPIs)

II.2.1 Classification

PPIs occur by means of hydrophobic binding, van der Waals forces and salt bridges, which occur in specific binding domains of each protein. There are different ways to classify PPIs based on their structural and functional characteristics[11].

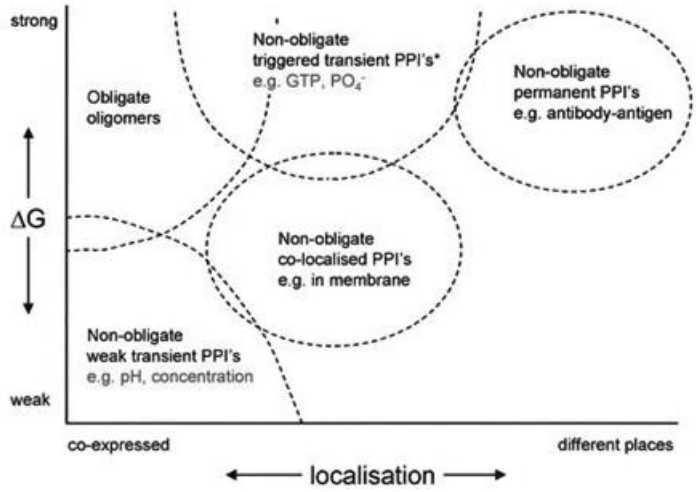


Figure 16 The relation between different types of protein–protein interactions (PPIs), their binding energy and the localization of their promoters.
Figure taken from [11].

PPIs can be classified based on their composition, affinity and life-time as homo- or hetero-oligomeric complexes, non-obligate or obligate complexes and transient or permanent complexes. Depending on their persistence, PPIs can be stable (permanent) or transient. In general, stable interactions are characteristic of protein complexes, which can include the same or different subunits. The hemoglobin complex is a representative example of a stable interaction between alike proteins. In contrast, most of the processes in a cell are a lot more dynamic and require transient interactions, which can quickly activate or stop a particular process. Those interactions can be either fast or slow. Two Src homology domains, SH2 and SH3, are examples of protein domains involved in common transient interactions. The binding domains of proteins can be a small cleft, a groove or a large surface. The binding interface will influence the strength of the binding. For example, leucine zipper domains, which display interaction surfaces, induce stable PPIs while SH2 and SH3 domain

interactions induce transient interactions. The difference between obligate and not obligate interactions depends on binding affinities. In an obligate interaction the components of the complex cannot exist independently due to their intrinsic instability whereas the constituents of a non-obligate interaction are stable on their own. An obligate homodimer is the DNA-binding P22 Arc repressor, while an obligate heterodimer is the human cathepsin DA. An example of a non-obligate homodimer is Sperm lysin, whereas the RhoA - RhoGAP signaling complex represents a non-obligate heterodimer. Of note, this is not a very strict classification. Every PPI may combine characteristics of these specific classifications while many PPIs do not strictly belong to a specific type. More specifically, physiological conditions and environments play a big role in influencing the characteristics of the interactions[11,12] (**Figure 17**).

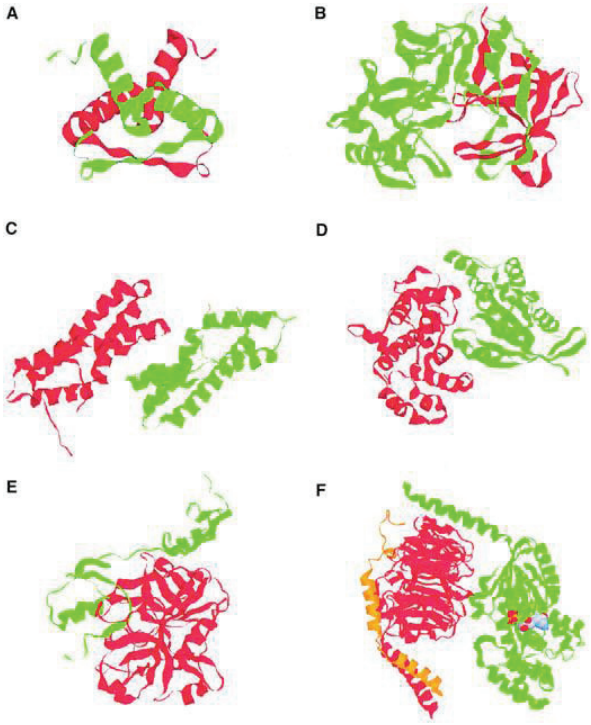


Figure 17 Examples of different types of protein–protein interactions
(A) Obligate homodimer, P22 Arc repressor; (B) Obligate heterodimer, human cathepsin D that consists of a non-homologous light (red) and heavy (green) chain; (C) Non-obligate homodimer, sperm lysine; (D) Non-obligate heterodimer, RhoA (green) and RhoGAP (red) signaling complex; (E) Non-obligate permanent heterodimer, thrombin (red) and rodniin inhibitor (green); (F) Non-obligate transient heterotrimer, bovine G protein, i.e., the interaction between Ga (green) and Gb (red, orange) is transient. The proteins in an obligate interaction are not found as stable structures on their own in vivo. Picture from [11].

II.2.2 Posttranslational Modifications (PTMs)

There are two strategies adopted by eukaryotes to expand the coding capacity of the genome: the first includes alternative mRNA splicing and alternative promoter usage at the transcriptional level, and the second comprises various posttranslational modifications. PTMs allow biochemical mechanisms important for transient cell signaling processes. They occur after protein translation events to obtain in a regulated way a mature active or inactive product. Via PTMs, amino acid residues are covalently modified. The site of the PTMs is the amino acid side chain or the C- or N-terminus of the modified protein. For instance, a clear correlation has been shown between cancer generation or progression and a high rate of mutations of protein PTM effectors (such as kinases, phosphatases, or ubiquitin E3 ligases)[13–16]. There are many different types of PTMs. One category of PTMs is covalent cleavage of the peptide, which is mainly achieved by proteases and occurs through a break of the peptide backbones. It is an irreversible and ubiquitous process that can result in activation/inactivation of the protein and regulation of biological processes[17]. A simple example of this type of PTM is the cleavage of the initiation methionine, which produces an active form of the protein[18]. A majority of PTMs is based on enzyme-mediated mechanisms adding electrophilic groups to nucleophilic side chain residues. The major PTMs of this type are phosphorylation, ubiquitination, methylation and acetylation. Phosphorylation is catalysed by kinases, which transfer a phosphate group from ATP to the side chain of serine (S), threonine (T) or tyrosine (Y) of the protein substrate[19–21]. These modifications are reversible by virtue of phosphatases[22]. Ubiquitination occurs when ubiquitin molecules are added to substrate proteins. The regulation of this modification is responsible for the dynamic state of synthesis and degradation of target proteins. Ubiquitin-protein ligases catalyse the process while deubiquitinating enzymes (DUBs) oppose the role of the ligases[23–26]. Methylation is another main PTM: it is a reversible process catalyzed by methyltransferases, which use lysine and arginine residues as the primary acceptors of methyl groups[27,28]. Acetylation is the result of attaching acetyl groups to the lysine side chains of proteins. A collaboration between acetylation events and other PTMs has been shown to play a role in the dynamic control of different signaling pathways[29–31]. PTMs are much more predominant in eukaryotes than prokaryotes, regarding both PTM type and frequency, even if recently many studies have shown important regulatory roles of protein modification in bacteria[32]. As a result of PTMs, different protein isoforms are generated, which may vary in properties and also in interacting partners. This is the reason why it is important to consider the proper system when PPIs are investigated. Allowing the protein under investigation to undergo the expected *in vivo* PTMs is expected to facilitate the identification of physiologically relevant interaction partners.

II.2.3 Methods to study PPI

Taking advantage from the different events occurring during the PPI process, many methods have been developed to study the way that proteins interact with each other. Each method is characterized by different levels of specificity and sensitivity and most of them are used at the same time to obtain a more complete or specific analysis. Genetic, biochemical, computational and biophysical technologies have been developed to study PPIs (a summary can be found in **Table 4**).

In this chapter, genetic and biochemical methods are presented and a quick overview of computational methods is given. The biophysical technologies (not presented here) have the advantage of giving information about the kinetics of the interactions. They use a range of biophysical characteristics, such as the changes in emission of heat or in speed of rotation that proteins undergo during their associations. Other biophysical specificities have been used to develop more techniques including dual polarization interferometry, surface plasmon resonance, static light scattering and circular dichroism methods[33].

Next to the *in vivo* and the *in vitro* methods to detect PPIs, which are overviewed in the following paragraphs, *in silico* methods have been developed to support experimental findings (complement the interactions detected by 'wet' experimental work). A list of those methods (together with the *in vivo* and *in vivo* ones) with a summary of their respective characteristics can be found in **Table 4**[12] .

Table 4 Summary of PPI detection methods (modified after Srinivasa rao et al., 2014[12]).
[(#), techniques used in this dissertation; (*), techniques described in this dissertation]

Approach	Technique	Summary
In vitro	Tandem affinity purification-mass spectroscopy (TAP-MS) (*)	TAP-MS is based on the double tagging of the protein of interest, followed by a two-step purification process and mass spectroscopic analysis
	Affinity chromatography (*)	Affinity chromatography is highly sensitive, can even detect weakest interactions in proteins, and also tests all the sample proteins equally for interaction.
	Coimmunoprecipitation (#) (*)	Coimmunoprecipitation confirms interactions using a whole cell extract where proteins are present in their native form in a complex mixture of cellular components
	Protein microarrays (*)	Microarray-based analysis allows the simultaneous analysis of thousands of parameters within a single experiment
	Phage display	The phage-display approach is based on the incorporation of the protein and genetic components into a single phage particle
	X-ray crystallography	X-ray crystallography enables visualization of protein structures at the atomic level and enhances the understanding of protein interaction and function
	NMR spectroscopy	NMR spectroscopy can even detect weak protein-protein interactions

In vivo	Yeast 2 hybrid (Y2H) (#)(*)	Yeast two-hybrid is typically carried out by screening a protein of interest against a random library of potential protein partners
	Synthetic lethality	Synthetic lethality is based on functional interactions rather than physical interaction
	MAPPIT (#)(*)	A cytokine receptor-based two-hybrid system that operates in mammalian cells. Occurring at the plasma membrane and is based on JAK-STAT signal transduction pathway.
	KISS(*)	KInase Substrate Sensor method extending the MAPPIT technology to a wider range of PPIs including membrane associated PPIs.
	Protein-fragment complementation (*)	Protein-fragment complementation assays (PCAs) can be used to detect PPI between proteins of any molecular weight and expressed at their endogenous levels
	FRET/BRET (*)	Based on RET are used as the common tools in the study of biochemical reaction kinetics.
In silico	Orthologue-based sequence approach	Orthologue-based sequence approach based on the homologous nature of the query protein in the annotated protein databases using pairwise local sequence algorithm
	Domain-pairs-based sequence approach	Domain-pairs-based approach predicts protein interactions based on domain-domain interactions
	Structure-based approaches	Structure-based approaches predict a PPI based on structural similarities (primary, secondary, or tertiary)
	Gene neighborhood	If the gene neighborhood is conserved across multiple genomes, then there is a potential possibility of the functional linkage among the proteins encoded by the related genes
	Gene fusion	Gene fusion, which is often referred to as Rosetta stone method, is based on the concept that some of the single-domain containing proteins in one organism can fuse to form a multi-domain protein in other organisms
	In silico 2 hybrid (I2H)	The I2H method is based on the assumption that interacting proteins should undergo coevolution in order to keep the protein function reliable
	Phylogenetic tree	The phylogenetic tree method predicts the protein-protein interaction based on the evolution history of the protein
	Phylogenetic profile	The phylogenetic profile predicts the interaction between two proteins if they share the same phylogenetic profile
	Gene expression	Gene expression may predict interactions based on the idea that proteins from the genes belonging to the common expression-profiling clusters are more likely to interact with each other than proteins from the genes belonging to different clusters

II.2.3.1 Genetic methods

II.2.3.1.1 Yeast Two-hybrid (Y2H)

General

Until 1989, PPIs have been studied solely by using biochemical techniques such as co-fractionation by chromatography, crosslinking, co-immunoprecipitation, which all require *in vitro* handling of protein extracts. The first genetic screen in yeast cells was described by Stanley Fields and Ok-kyu Song in 1989, and revolutionized protein interaction analysis[34]. This so-called Yeast Two-Hybrid (Y2H)

approach became one of the most widely used methods to investigate PPIs until today. Many related methods have been developed since then to answer different questions. In this chapter the Y2H approach and variant Y2H methodologies will be described.

The Y2H method

Only few years before the seminal publication by Fields and Song[34], the Ptashne laboratory had discovered the modular structure of GAL4, a transcriptional activator in yeast[35]. The yeast GAL4 transcription factor consists of separable domains responsible for DNA binding and transcriptional activation. These researchers showed that Gal4 binds the upstream activation element of target genes and thus activates transcription in the presence of galactose. They demonstrated that the N-terminal fragment alone can still bind to DNA, but cannot activate transcription in the presence of galactose. This means that GAL4 cannot activate transcription unless it is associated with an activation domain. Their findings demonstrated the role of the C-terminal fragment of GAL4 as an activation domain. The Y2H method takes then advantage of the fact that both fragments are in proximity, even when not covalently linked, and reconstitute a fully functional protein.

The Y2H method is thus based on the fact that most eukaryotic transcription factors show a modular structure composed of a DNA binding domain (DBD) linked to an activation domain (AD). When both domains are in proximity then a functional transcription factor can be obtained[36].

Technically, a cDNA sequence encoding a bait protein of interest is genetically fused to the DBD of a transcriptional activator and a cDNA sequence encoding a prey protein (fragment) is fused to the AD of the transcription factor. The interaction between the two-hybrid proteins reconstitutes the transcriptional activity. In this way, the DBD binds a specific region of the reporter gene promoter while the AD interacts with the RNA polymerase II, resulting in the activation of a reporter gene or selected genes. A measurable output is generated, which can be a reporter signal or growth under selection pressure (**Figure 18**).

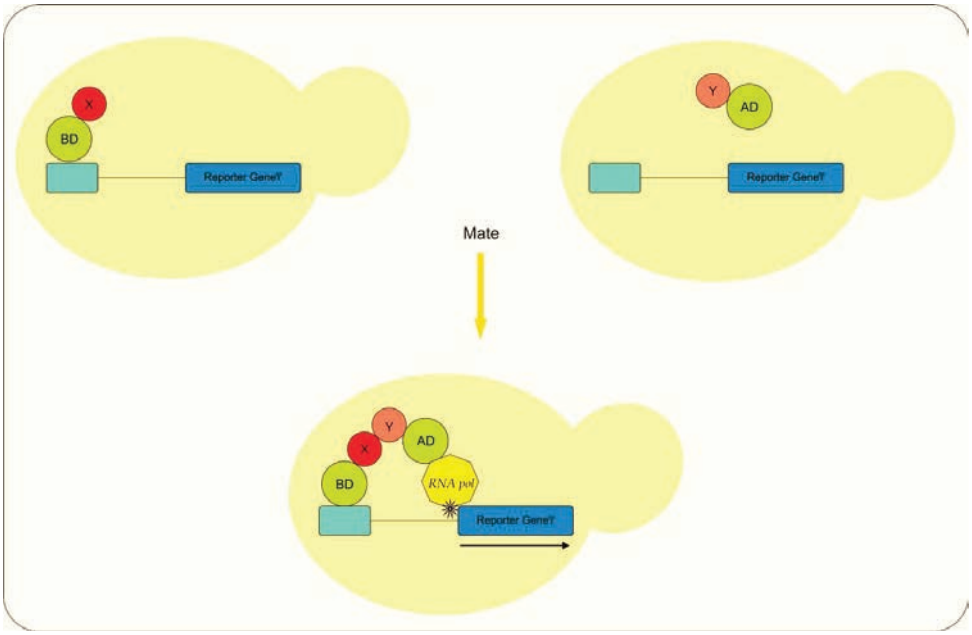


Figure 18 Y2H, the concept
In Yeast, haploid cells mate with other haploid cells of a different mating type (an "a" cell can only mate with a "α" cell but not with another "a" cell, and vice versa). After secreting pheromones and growing projections, the mating occurs and produces a stable diploid cell. The interaction of the bait (X) with the prey (Y) in the diploid cell leads to the reconstitution of a functional transcription factor by complementing its DBD and the AD. The AD interacts with RNA pol II driving transcription of the reporter gene.

In the original paper, Fields and Song used the yeast GAL4 transcriptional activator to confirm the known interaction between SNF1 and SNF4. This interaction led to the transcription of a *GAL1-lacZ* fusion gene. A yeast strain expressing both hybrid genes induced the expression of the reporter gene encoding the enzyme beta-galactosidase, which labels the yeast cell blue when using a colorimetric substrate[34].

Pros and cons of the Y2H approach

Thanks to several advantages, the Y2H approach is the most widely used technique to investigate PPIs. The system is highly sensitive hence weak or transient interactions can be detected[37]. Moreover, the technique is cheap, fast and simple requiring limited technical resources. It can be used to investigate specific PPIs but also for high-throughput screening procedures[38,39]. Furthermore, the method uses living cells and provides an eukaryotic environment. In contrast with the classic biochemical approaches, no production and purification of proteins is required.

Despite the high number of benefits, the technique is not perfect and shows limitations. A major problem of the method is the high percentage of false positives. False positives are physical interactions detected, which cannot be reproduced in a different system and might be generated due to technical (by an event different from PPI) or biological (PPI not occurring *in vivo*) reasons. Y2H experiments work through overexpression of prey and bait plasmids, which can result in nonspecific interactions. Moreover, the overexpression occurs in a compartment (nucleus), which does not necessarily correspond to the natural cellular context of the hybrid proteins and the use of fusion proteins implies that their conformations may differ from the native ones, compromising their binding abilities. Self-activation of the reporter gene individually by the bait and/or the prey proteins will result in a false positive signal too. Another problem of the Y2H system is the generation of false negatives, defined as PPIs which cannot be detected due to limitations of the method. For example, interactions including membrane proteins are mostly undetectable in Y2H. Furthermore, interactions of proteins of higher eukaryotes may not be detected due to the absence of posttranslational modifications, which may be hard to reproduce correctly in the lower eukaryote yeast. Between two independent large-scale Y2H screens, less than 30% of the identified interactions overlap[39]. This is caused by a low assay sensitivity, which implies a large number of not identified true interactions, but also by the low coverage which depends from the used collection and low sample sensitivity, defined as the number of detected interactions in a single trial. Already in the 1989, Fields and Song were aware of the following limitations of their method: "The system requires that the interaction can occur within the yeast nucleus, that the Gal4-activating region is accessible to the transcription machinery and that the Gal4 (1-147)-protein X hybrid is itself not a potent activator". Due to these severe limitations the sensitivity of the Y2H is estimated to ~25%, thus there is a growing interest for the development of alternative Y2H variants[40].

II.2.3.1.2 Variant Y2H methodologies

One of the most obvious limitations of the Y2H methods is the impossibility of studying proteins that can directly activate the transcription, independently of their fusion context as a prey protein. To solve this problem two different methods have been developed.

First, the repressed trans-activator (RTA) system was introduced[41] (**Figure 19A**). Here, while the bait fusion protein is able to activate transcription, the prey proteins are fused to the N-repression domain (RD) of a transcription repressor (e.g. Tup1). Interaction between the activator bait and a RD prey fusion causes repression instead of activation of the reporter gene. This strategy has been used

to identify novel protein interactions with a variety of transcriptional activators, including herpes simplex virus 1 (HSV-1) regulatory protein VP16[41] , c-myc [42], and the androgen receptor[43].

Another variation on the original model is the RNA polymerase III-based two-hybrid system, which was developed to detect interactions that cannot be studied by the original RNA polymerase II two-hybrid system (**Figure 19A**). Here the prey protein is combined with a different fusion partner, namely a subunit of the multimeric protein complex TFIIIC, which is one of the two transcription factors involved in RNA polymerase-III mediated transcription. This makes the method RNA polymerase-II independent, and avoids false positivity due to trans activation by the bait fusion protein.

The original Y2H method and the two variants presented above are based on the translocation of the hybrid proteins into the nucleus, which makes the method unsuitable for a high percentage of protein interactions. This is the reason why different approaches have been developed where the interacting cytoplasmic proteins or protein domains are in their natural cellular compartment.

One method to analyze cytosolic proteins, later extended to membrane-bound proteins, is the ubiquitin-based split protein sensor[44] (**Figure 19D**). Ubiquitin (Ub) is a 76 amino acid protein important for the natural turnover of cellular proteins. Proteins, to which ubiquitins are covalently bound, are labelled to undergo proteasomal degradation. The split ubiquitin method is based on the separation of ubiquitin into two fragments, the C-terminal half (Cub) and the N-terminal half (Nub) including a mutation (I13G or I13A) to avoid spontaneous dimerization. If interaction between the fragments occurs, it results in a reconstituted split-Ub. The cleavage by an ubiquitin-specific protease releases a transcription factor and activates the reporter gene expression.

The SOS- and RAS-recruitment systems are two other variants developed to study cytosolic and membrane proteins (**Figure 19B**). They don't need a transcriptional read out since they use the Ras signaling pathway, which is homologous between yeast and mammals[45]. Ras is a G protein, also called small GTPase, which means that it is a guanosine-nucleotide-binding protein that functions as a binary signaling switch with "on" (GTP-bound) and "off" (GDP-bound) states. The switch happens at the plasma membrane and is mediated by GEF proteins, Sdc25p in yeast and SOS in mammals. The system takes advantage of the capacity of hSOS to substitute the yeast Ras GEF (guanine exchange factor), Cdc25 in a cell with mutant cdc25 gene.

In the SOS recruitment system the bait is fused to mammalian SOS; if the bait interacts with a prey in the proximity of the plasma membrane, SOS activates the GDP to GTP exchange in Ras, activating in

this way the signaling pathway. In the Ras recruitment system (RRS), the bait protein is fused to a constitutively active mammalian Ras. In this case, only the correct localization is needed to activate the downstream signaling. The interaction with a membrane prey will mediate this translocation. The activation of the signal cascade causes cell survival and growth.

For the analysis of membrane-localized PPIs, a variation of the RRS has been developed, which has been called reverse RRS (rRRS) Y2H. The bait is here an intrinsic membrane protein or a membrane-anchored protein and the prey is the Ras fusion protein. In 2000, Ehrhard and colleagues, developed a similar membrane-associated two-hybrid assay based on the recruitment of a chimeric G protein subunit to the membrane due to the interaction with a Ste2 hybrid protein that leads to the signaling cascade activation[46].

In the past decades, the Y2H system has been adapted to cover an increasingly wide range of applications, including the targeting of PPIs for therapeutic actions. Y2H variants have been developed to investigate interaction of proteins and small molecules in the best possible physiological setting[47]. First, the so-called reverse Y2H can detect mutations or molecules that dissociate PPIs. To this end, Vidal *et al.*, used the gene URA3 as a reporter[48]. Expression of URA3 is toxic for cells growing in a medium containing 5-fluoroortic acid, therefore the interaction of the proteins of interest will lead to cell death. The inhibition of the association would thus lead to cell growth. Another variation is the yeast three hybrid that allows the screening for proteins that directly interact with a specific molecule *in vivo*[49]. In this method, the third molecule is a synthetic heterodimer of two different small organic ligands.

For further details on different Y2H system we refer to **Table 5** and **Figure 19**[40,47].

Table 5 Overview of different Y2H systems

Y2H Method	Possible bait	Response	Principle	Reference
Classic Y2H System	Non-trans activating proteins capable of entering the nucleus	Transcriptional Activation	Artificial recruitment of bait and prey proteins to the nucleus	[34]
SOS recruitment system	Trans activating, cytosolic proteins	Ras signaling	Detection of PPIs at the membrane on the basis of the defective cdc25 allele in yeast. A PPI translocates a Ras fusion protein to the membrane and allows growth	[50–52]
Split Ubiquitin system (MYTH)	Nuclear, membrane and cytosolic proteins	Uracil autotrophy and 5-FoA resistance	The reunion of two ubiquitin fragments in the cytosol due to a PPI establishes the cleavage activity which releases a transcription factor allowing activation of the reporter gene.	[53,54]
RAS recruitment system	Trans activating, cytosolic proteins	Ras Signaling	See "SOS recruitment system"	[45,50,51]
Dual bait system	Two non-trans activating proteins capable of entering the nucleus	Transcriptional Activation	Y2H with two baits. Interaction occurs In the nucleus.	[55]
G-protein fusion system	Membrane protein	Inhibition of protein G signaling	To monitor PPIs between integral membrane proteins.	[56]
RNA polymerase III based two-hybrid	Trans activating proteins in the RNA polymerase II pathway	Transcriptional Activation	Activation of RNA-polIII transcription by recruitment of transcriptional activator	[57]

Repressed trans activator system	Trans activating proteins capable of entering nucleus	Inhibition of transcriptional activation	Interaction between transcriptional activator in the nucleus.	[41,43]
Reverse Ras recruitment system	Membrane proteins	Ras signaling	The prey is a Ras fused protein and the bait a membrane protein. Principle of the RRS	[58]
CINEX-P System	Extracellular proteins	Downstream signaling and transcriptional activation	System to show PPIs in the ER based on the Ire1p, a type 1 ER membrane protein and the activation of the unfolded protein response	[59]
Split-Trp System	Cytosolic, membrane proteins	Trp1p activity	The split-Trp protein sensors allow for the detection of PPIs by enabling trp1 cells to grow on medium lacking tryptophan	[60]
Cytosolic split-ubiquitin system	Trans activating, cytosolic proteins	Transcriptional activation	Used for membrane anchored cytosolic baits and occurs close to the ER membrane	[61]

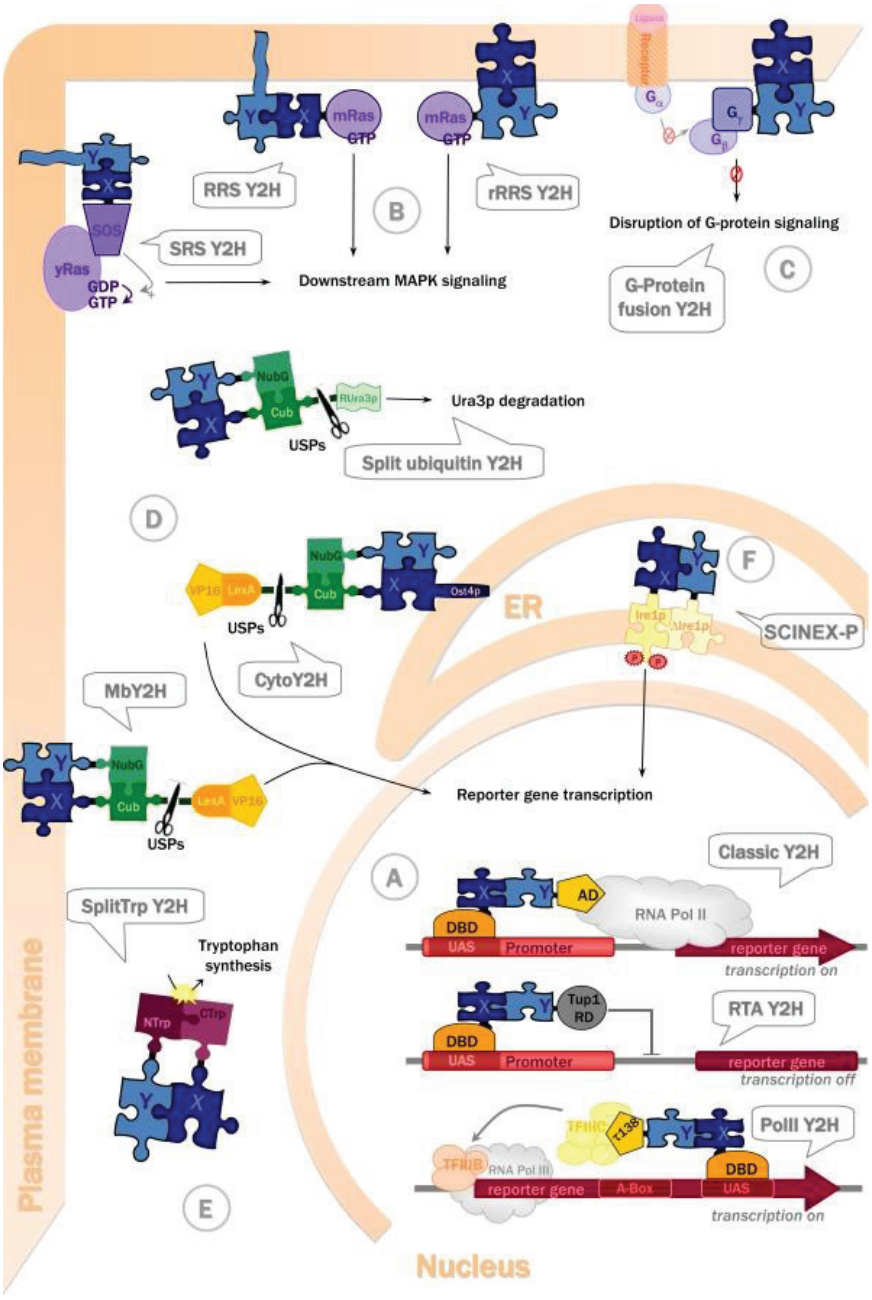


Figure 19 The Yeast two-hybrid systems, their subcellular location within a yeast cell, and their operating mode (represented at the moment of bait-prey interaction).

Protein X: bait, dark blue puzzle piece; Protein Y: prey, light blue puzzle piece.
When X and Y interact three events can occur: reconstitution of a split-protein, like in A, D and E; membrane recruitment, such as in B and C, or protein dimerization, F.
(A) Nuclear Y2H systems occur in the nucleus. Y2H and RTA Y2H require RNA pol II activity while PolIII Y2H involves RNA pol III. (B) Ras signaling based Y2H: SRS, RRS and rRRS Y2H are based on the recruitment of the proteins to the membrane. The interaction of X and Y activate the MAPK downstream pathway. (C) G-protein signaling-based Y2H: able to investigate interaction between integral membrane bait and a soluble prey fused to the γ-subunit of a heterotrimeric G-protein. The interaction of X and Y lead to the recruitment of G-protein β-subunits and the subsequent inactivation of the G protein signaling pathway, involved in pheromone response and used as a read out. (D) Split-ubiquitin based Y2H systems: the subcellular localization depends from the nature of X and Y and on the reporter. The interaction leads to complementation of ubiquitin followed by proteolytic release of the dha reporter detected via western blot, or of a transcription factor or of a destabilized version of the URA3 enzyme. RUra3p undergo rapid degradation after being released. In absence if the enzyme cells grow in medium with 5-FOA and this is used as read out. The MbY2H (membrane trans activator split-ubiquitin) system is used for interaction between membrane proteins of the endoplasmic reticulum. The CytoY2H is an adaptation of MbY2H to investigate cytosolic proteins which are anchored to the ER membrane. (E) Split-protein sensor Y2H. The Split-Trp Y2H is based on the reconstitution of the activity of an enzyme responsible for the synthesis of tryptophan. (F) ER Y2H system. The SCINEX-P (screening for interactions between extracellular proteins) system is based on signaling of the yeast unfolded protein response (UPR) using the properties of Ire1p, a type I transmembrane (TM) protein that resides in the yeast E. The activation of Ire1p leads to translation of HAC1. In the system, X is fused to an Ire1 variant (Ire1p) able to splice Hac1 mRNA for translation only when the interaction with Y occurs. The prey is fused to a variant of ire (ΔIrep) able to phosphorylate only the associated Ire1 partner. The phosphorylation activates Ire1p and the transcription of the reporter gene. (Figure from[40]).

II.2.3.1.3 Genetic Protein-Protein Interaction Methods in Prokaryotes: a quick overview

Despite the limitation given by the lack of PTMs, prokaryotes are still considered a valid alternative to yeast for two-hybrid methods. Working with *Escherichia coli* allows very fast experiments given the fast growth of the bacteria and the high transformation efficiency. The sensitivity of a prokaryote system is higher if we consider the lack of endogenous proteins, which could compete with bait and pray for binding. Additionally the toxicity levels in yeast can differ from prokaryotic cells. Different systems have been developed in *E. coli* for PPIs studies[33].

The Repressor-based two-hybrid system was the first bacterial two-hybrid approach to be developed[62]. It uses the LexA repressor and is able to detect interactions via a technically easy and quantitative assay. The repressor is a homodimeric DNA binding protein where the N-terminal domain represents the DNA binding domain (DBD) and the C-terminal domain is responsible for the dimerization necessary for stable DNA binding. The proteins of interest X and Y can be fused each with a different variant of the LexA repressor to avoid prey interactions. If they interact, the heterodimerization induces an efficient repression of the reporter gene lacZ. Quantitative β-galactosidase assays are performed to measure the effect of the two-hybrid proteins on reporter gene expression. An adaptation of the technique was developed by the use of the bacteriophage λ cI

repressor as DBD and the subunit of RNA polymerase as AD[63]. Joung and colleagues[64] developed another variant where a motif of murine Zif268 is used as DBD and the AD is an operon of HIS3. The output is given by the activity of the spectinomycin resistance gene aadA. Other variants of the original system are listed in **Table 6**, shown in **Figure 20** and reviewed by Stynen *et al.* [33].

Table 6 **List and description of Y2H variants**

Method Variants	Description And References		Figure 20
Repressor-based two-hybrid system	Interaction between two proteins of interest, X and Y, can be monitored by fusion of each with a variant of the LexA repressor (408 or wild type). Heterodimerization given by the interaction activate the repression of the reporter lacZ.	[62]	(A)
PhaR two-hybrid system	The bait X is fused to the DNA-binding domain (DBD) of the repressor PhaR and the prey Y is fused to PHB granule-associated protein PhaP: the interaction permit the recruitment of PhaR to PHB granuls which results in lacZ expression.	[65]	(B)
GFP recruitment system	The bait X localizes at the cell division sites thanks to the activity of fusion protein *B. subtilis* DivIVA. The prey protein is fused to GFP. In case of interaction, a fluorescence will be localized the division sites. (GFP recruitment systems are also available for the pathogenic fungus Candida albicans, for C. elegans, and for mammalian cells.)	[66]	(C)
ToxR two-hybrid system	The *V. cholerae* ToxR transcriptional activator requires dimerization of its periplasmic domain for full reporter transcriptional activation. In the ToxR two-hybrid system, the periplasmic domain is replaced by two proteins of interest, X and Y. An interaction between these two proteins results in efficient ctx promoter binding of the truncated ToxR protein (ToxR=) and subsequent gene expression of lacZ or the chloramphenicol acetyltransferase gene (cat)	[67]	(D)
Tat two-hybrid system	The Tat signal sequence (ss) peptide tethers bait protein X to the periplasm. A chimeric fusion of prey Y with the maltose-binding protein without a signal sequence (ssMBP) localizes to the periplasm only upon interaction of X with Y. This translocation is required for growth on medium with maltose as the sole carbon source. Alternatively, the prey protein Y is fused to a localization-deficient DsbA enzyme (ssDsbA), which catalyzes the formation of active alkaline phosphatase (AP). Active AP converts p-nitrophenyl phosphate (pNPP) to yellow p-nitrophenol (pNP)	[68]	(E)
Bacterial two-hybrid system for DNA-protein interactions	To increase sensitivity in the search for zinc finger-DNA associations, the binding of a zinc finger motif X to its target DNA sequence, Y (zinc finger binding motif Zfbm Y), is facilitated by inclusion of two fixed zinc fingers from Zif268 and the target DNA sequence Zif268 bm. The zinc finger fusion further consists of the S. cerevisiae Gal11 interaction domain (Gal11 ID), which binds the S. cerevisiae Gal4 dimerization domain (Gal4 ID). The latter domain is fused to the N-terminal domain of the RNA polymerase subunit for indirect activation of an operon comprising the S. cerevisiae auxotrophic marker HIS3 and aadA, which	[64]	(F)

	confers resistance to spectinomycin.		
Bacterial reverse two-hybrid system	Interaction between chimeric proteins of bait X with the bacteriophage cI repressor (which binds the cI OR2 operator) and prey Y with the N-terminal domain of the RNA polymerase subunit results in activation of a gene encoding the C-terminal domain of the bacteriophage 186 cI repressor (186 cI CTD). This truncated protein sequesters and inactivates full-length 186 cI, which normally downregulates cytotoxic 186 prophage genes. Resulting cell death can be circumvented by mutations that block the bait-prey interaction.	[69]	(G)
Intein-mediated split-GFP assay	Bait protein X is in fusion with the N-terminal fragments of the intein VDE and GFP, while prey protein Y constructs include their respective C-terminal counterparts. Interaction between X and Y reconstitutes VDE, which splices out and covalently reattaches the GFP fragments to create an isolated GFP monomer, detected by fluorescence.	[70]	(H)

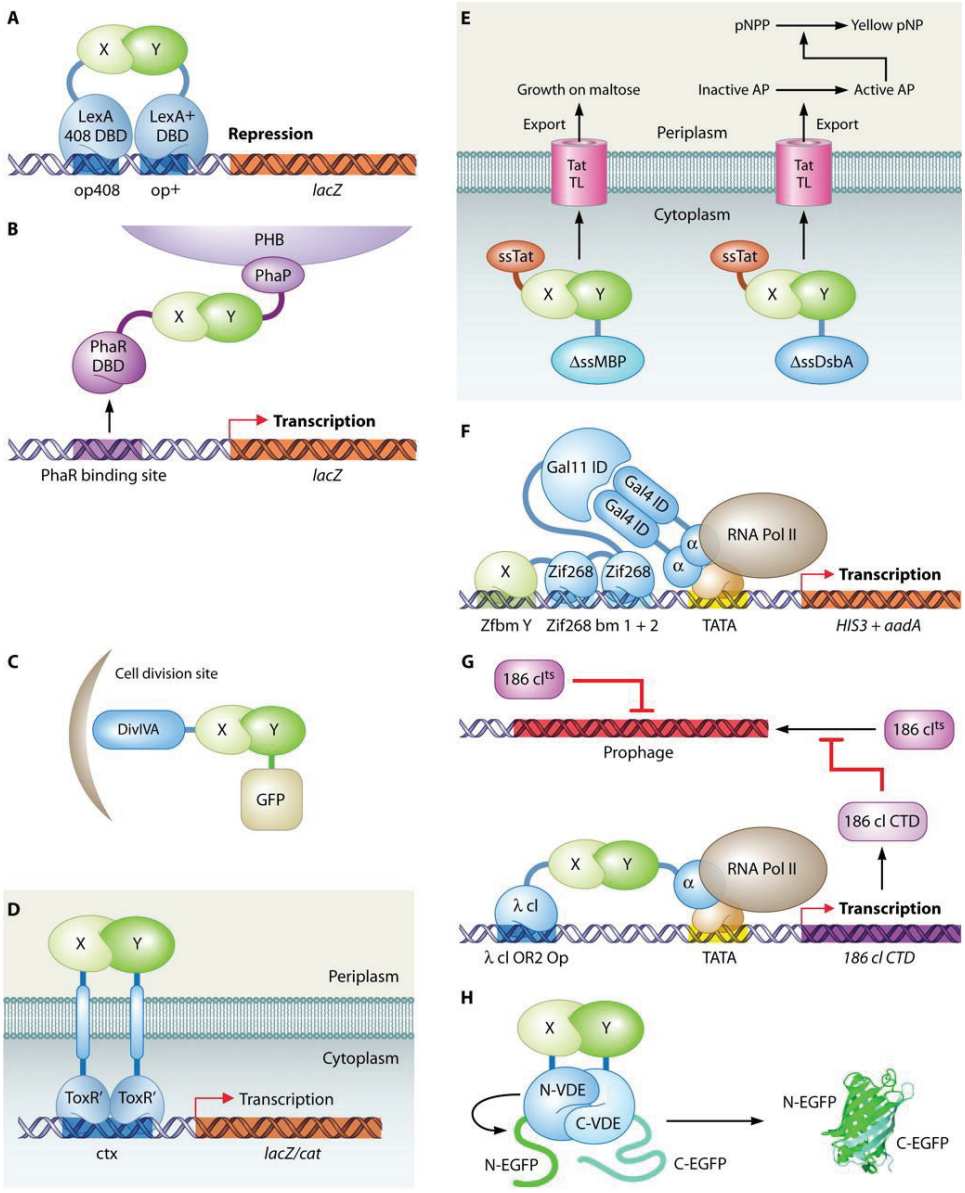


Figure 20 Specific bacterial genetic PPI detection methods.

(A) Repressor-based two-hybrid system (B) PhaR two-hybrid system. (C) GFP recruitment system. (D) ToxR two-hybrid system. (E) Tat two-hybrid system. (F) Bacterial two-hybrid system for DNA-protein interactions. (G) Bacterial reverse two-hybrid system. (H) Intein-mediated split-GFP assay. For references and details see text, table and Stynen et al., 2012[33]. (Figure and description adapted from[33])

II.2.3.1.4 Two-hybrid systems in mammalian cells

As described above, the Y2H system shows several limitations despite its high number of benefits. The mammalian two-hybrid systems have been developed to study human or mouse protein interactions in a more natural environment. These systems may better mimic the *in vivo* situation, in terms of the natural folding, the posttranslational modifications and the cellular localization of the proteins being investigated.

M2H system

The mammalian two-hybrid system (M2H system) is a conceptual replica of the Y2H: it is a genetic, *in vivo* assay based on the reconstitution of the function of a transcriptional activator. The GAL4 DBD is fused with protein X and the AD of the Herpes simplex VP16 protein is coupled to protein Y. The interaction of protein X with protein Y activates the transcription of the reporter gene. As reporter genes are mostly used those encoding luciferase, beta-galactosidase or chloramphenicol acetyltransferase (CAT)[71]. The M2H system can be used to verify PPIs detected via Y2H screenings and it was first used to study leucine zipper interactions: in 1991 Dang *et al.* observed, via M2H, the oligo dimerization in Chinese Hamster Ovary cells between leucine zipper regions from c-Jun and c-Fos[72].

Based on the same principle of the Ras recruitment system in yeast, an adaptation of the system was also developed in mammalian cells, the so-called mammalian RRS: it can be used as a validation method of the yeast RRS system basically for every cytoplasmic protein (except for the ones interfering with the MAPK pathway)[73]. This method was used to confirm the interaction between Grb2 and Sos in HEK293 cell line and p11 and p85 of the phosphatidylinositol 3-kinase (PI3K)[73].

Protein fragment complementation assays (PCA)

Protein fragment complementation assays (PCAs) are another set of techniques that allow the detection of PPIs in cells *in vivo* [74]. PCAs are based on the principle that proteins should exist in close proximity so that complementation and folding can occur (different PCA strategies can be found in **Table 7**). From the ubiquitin-based split protein sensor, which is based on the complementation of two inactive subunits to form a functional protein in yeast (see paragraph II.2.3.1.2 of this dissertation), different analogous techniques have been developed. The main differences are in the

type of protein fragments. Briefly, a functional protein is reconstituted by the interaction between bait and prey tagged with non-functional halves of this protein (**Figure 21**).

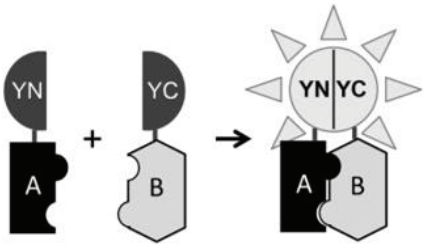


Figure 21 Schematic representation of the BiFC analysis. The N-terminal and C-terminal fragments (YN and YC) of YFP are fused to two proteins of interest (A and B). If A and B interact, the formation of a fluorescent complex occurs. Modified from [75,76].

One of the most important advantages of the technique is that the interaction generates a functional protein that can activate the reporter gene or even act as a reporter itself, providing the read out of the assay directly where the interaction occurs[77,78]. Many reporter proteins can be used and different PCA strategies have been developed using the complementation of diverse proteins as a read out. An overview of different PCA strategies is shown in table 6. When the strategy is applied in the context of fluorescent proteins, the technique is named bimolecular fluorescence complementation assay or BiFC. Originally, the Fluorescence complementation (FC) was shown *in vitro* using GFP fragments in *Escherichia coli*[79]. This was the basis for Kerppola and colleagues to develop the FC method in mammalian cells, using a yellow fluorescent protein (YFP): it was first used for the visualization of the interactions between the basic leucine zipper (bZIP) and NF-κB family of proteins in mammalian cells. The unique characteristic of the method is that allows direct visualization of PPIs in living cells: the signal is directly visualized by fluorescent microscopy or analyzed by flow cytometry[76,80,81].

Table 7 PCA strategies, modified from[82]

Strategy	Readout System	Advantages	Disadvantage	Ref.
Split Ubiquitin System	Coupled reporter	Flexibility in reporter system. Signal amplification. No exogenous substrate for detection of complementation.	Require endogenous activity. Irreversible cleavage of reporter.	[83]
Split β-galactosidase assay	Absorbance Fluorescence	Signal amplification.	Irreversible complementation. Exogenous substrate required for detection of complementation.	[84]
Dihydrofolate reductase complementation	Fluorescence Cell survival	Signal amplification.	Irreversible complementation. Exogenous substrate required for detection of complementation.	[85]
β-lactamase protein fragment complementation	Absorbance Fluorescence	Signal amplification. Fast complementation.	Irreversible complementation. Exogenous substrate required for detection of complementation.	[86]
Bimolecular fluorescence complementation **(YFP fragments)**	fluorescence	High spatial resolution. No exogenous substrate for detection of complementation.	Slow fragment complementation. Irreversible complementation. No signal amplification.	[87,88]
Luciferase complementation	Bioluminescence	Fast complementation. Study of protein interaction dynamics via reversible complementation. Signal amplification.	Exogenous substrate required for detection of complementation.	[89]
Split TEV assay	Coupled reporter	Flexibility in reporter system. Signal amplification. No exogenous substrate required for detection of complementation.	Irreversible cleavage of reporter.	[89]

The mammalian-membrane two-hybrid assay (MaMTH)

MaMTH is based on the ubiquitin-based split protein sensor technique and it was developed for the study of interactions between membrane proteins in human cells in a stimulus-dependent manner. Ubiquitin is divided in two fragments, Nub and Cub. The prey protein is fused to the Nub fragment and the bait forms a fusion protein with the Cub fragment linked to a transcription factor (TF), which is cleaved from specific DUBs if interaction with the bait occurs. The TF migrates to the nucleus and activates the reporter gene[90] (**Figure 22**). As cell line, Hek293T cells stably expressing reporter constructs are used. The system was first successfully applied to show the known interactions between the ErbB4 and EGF receptors.

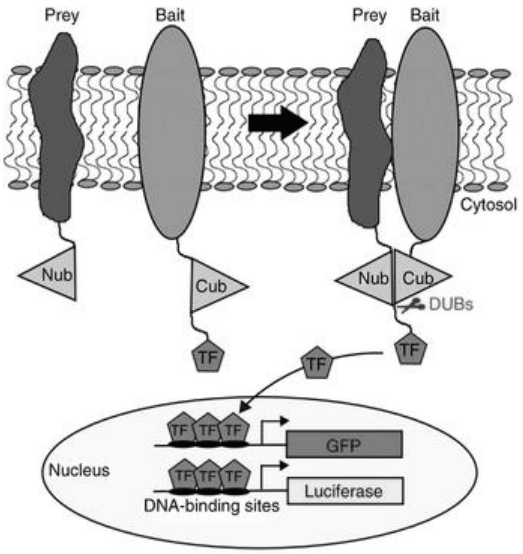


Figure 22 The MaMTH system.
In MaMTH bait and prey fusion proteins are cotransfected into HEK293T cells. In case of interaction, a TF is released that activates the reporter genes responsible for the transcription of luciferase or GFP.
Modified from [90].

Resonance energy transfer system

The resonance energy transfer (RET) principle[91] is based on Stryet's studies. He observed that electron excitation energy could be transferred over distances of the order of 30 Å. This process is the cause of light-induced damage, which occurs during the so-called photodynamic action[92].

Based on RET, two techniques have been developed in order to study PPIs in living cells: Förster resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET). The FRET is based on a radiation-free energy transfer between neighboring molecules. With conventional fluorescent microscopy it is possible to analyze the localization of labelled molecules in a spatial resolution limit of approximately 200 nanometers (0.2 micrometers). In order to study the occurrence of PPIs and the relative proximity between these proteins, new and not diffraction

limited approaches were required: the FRET technique, when applied to optical microscopy, is able to determine intermolecular distances of a few nanometers.

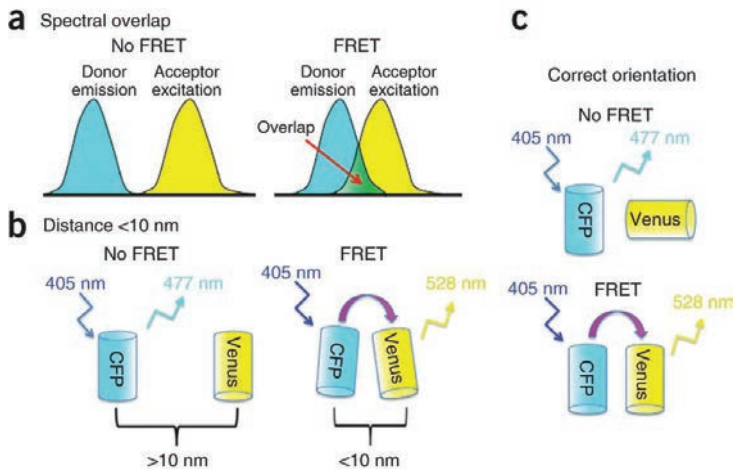


Figure 23 Schematic diagrams showing the three conditions necessary for efficient FRET.
Picture taken from [93]

The core of the technique is the fact that the energy transfer can occur between any two chromophores, assuming that the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor (**Figure 23 a**). Importantly, Förster and others have shown that the energy transfer efficiency is highly dependent on the distance between the donor and acceptor and their relative orientation (**Figure 23 b and c**)[92,94] .

Different fluorescent proteins can be coupled to the investigated proteins, characterized by distinct emission and excitation spectra, including cyan fluorescent protein (CFP), GFP, YFP and Dansyl-Fluorescein (FITC)[95].

The FRET technique opened possibilities for different applications. The intermolecular FRET (**Figure 24**) can occur between one

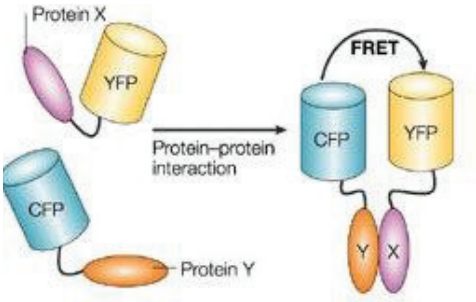


Figure 24 Intermolecular FRET
Description in the text. Figure modified from [96]

molecule (protein A) fused to the donor (CFP) and another molecule (protein B) fused to the acceptor (YFP). When the two conjugated proteins bind to each other, FRET occurs; when they dissociate, FRET diminishes.

The detection method of FRET is also a discriminant to obtain more accurate results: commonly the detection methods are based on intensity or fluorescence lifetime. Fluorescence lifetime imaging microscopy is considered the most accurate detection method[95]. For long time, the cyan and the yellow pairs of fluorescent protein have been the most common, but lately they have been replaced by green-red variants to permit a longer wavelength excitation and to reduce auto fluorescence and photo toxicity[96]. Despite the many advantages of using fluorescent proteins, it is important to mention the complications by auto-fluorescence, photo-bleaching or photochemical destruction of the fluorophores. To avoid this downside of the FRET technology, bioluminescence resonance energy transfer (BRET) has been developed. Instead of using a fluorescent protein as a donor, BRET utilizes a bioluminescent luciferase (typically from *Renilla reniformis)*. The donor and the investigated protein are fused while a GFP mutant is fused to the putative interaction partner. Luciferase, in the presence of its substrate (coelenterazine), emits bioluminescence leading to the excitation of the acceptor and consequently the emission of fluorescence. The transfer is possible if the distance between the luciferase and the fluorescent protein is less than 100 Å. The first example of BRET was the demonstration of the homodimerization of the clock protein KaiB from cyanobacteria[97].

MAPPIT

(Forward) MAPPIT is a cytokine receptor-based two-hybrid system that operates in mammalian cells[98]. The technique is based on type-I cytokine signal transduction. Type-I cytokine receptors such as erythropoietin receptor (EpoR), growth hormone receptor (GHR), leptin receptor (LR) and granulocyte colony stimulating factor receptor (G-CSFR) show a highly conserved structure in their extracellular domain, which is responsible for the binding of the ligand and they mostly

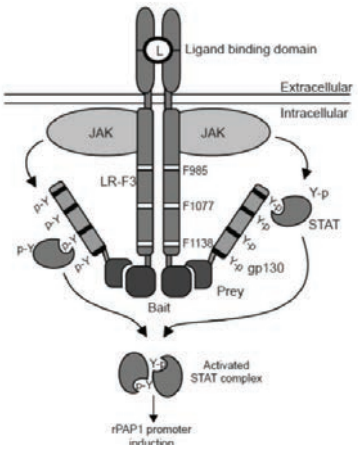


Figure 25 MAPPIT. The concept

transduce the signal intracellularly via the JAK-STAT pathway. Such receptors do not show any intrinsic kinase activity and phosphorylation occurs in association with JAKs (Janus kinases). Upon binding of the ligand to the receptor, a reorganization of the receptor subunits occurs which leads to the cross-phosphorylation of tyrosine residues on the receptor-associated JAK kinases. These events enable the recruitment and activation of STAT transcription proteins. Phosphorylated STATs migrate as dimers in the nucleus and activate specific gene transcription (**Figure 25** and **Figure 26a**). It has been shown in different receptor systems that the mutation of certain tyrosine residues in the cytosolic domain of the receptor leads to the complete loss of STAT activation. The MAPPIT technology applies this finding to create a non-functional mutated receptor able to restore its function in case of interaction between the bait and the prey.

The first hybrid protein used in MAPPIT is a receptor chain that consists of the extracellular domain of the human erythropoietin receptor (Epo) receptor and the transmembrane and intracellular domain of a mutated murine leptin receptor (LR) that still allows JAK activation but is deficient on STAT3 recruitment sites: tyrosine (Y) Y985, Y1077 and Y1138 are replaced by phenylalanine (F) residues. The mutation of the three conserved Y residues in its cytoplasmic domain to F knocks out STAT3 activation and also disables feedback mechanisms. The bait is C-terminally fused to such mutated receptor. The second hybrid used is the prey protein linked to a receptor fragment containing functional STAT3 recruitment sites, such as the C-terminal part of gp130, a signaling subunit of receptor complexes for ligands of the interleukin-6 family. After co-expression, bait and prey interactions are observed as functional complementation of the signaling pathway: upon stimulation with the ligand and bait-prey interaction, STAT3 molecules are able to migrate to the nucleus and induce transcription, as detected by using reporter or selection target genes. In the original model, the read-out was either a STAT3 phosphorylation test or a STAT3-based reporter assay leading to the expression of luciferase or an IL5Rα-derived membrane-tag. In the Binary MAPPIT models, the luciferase-based reporter assay is usually the read out of choice. For the detection, the reporter construct used consists of the luciferase gene under the control of the rat pancreatitis-associated protein 1 (rPAP1) promoter, which is a strong STAT3 responsive promoter[99]. The approach was evaluated using well-known interactions namely those between SV40 large T antigen (SVT) and p53, and between a member of the SOCS family and its binding motif (phosphorylated Tyr 402) in the Epo receptor, demonstrating how the technique can detect both modification–independent and tyrosine-phosphorylation dependent interactions[98]. Like every technique, MAPPIT technology has some limitations when it comes to general applications.

However, the derivation of related technologies from the standard protocol with different modes of action has enabled the analysis of a broad set of PPIs.

The MAPPIT system has been adapted to different needs. For example, to study protein interactions with the LR in its normal oligomeric configuration, different residues of Y can be mutated into F. In this case, LR itself is the bait protein and if the Y1138F mutation is present, STAT3 is not recruited. Upon stimulation, the other two tyrosine residues can be phosphorylated by JAK2. The interaction with the prey activates STAT3 and induces the reporter. The cytosolic domain of the receptor itself can be replaced by a flexible linker composed of Gly-Gly-Ser (GGS) repeats preventing any background activation resulting from prey association with the LR-F3. This replacement is possible since only the membrane proximal region of the LR is responsible for the binding with JAK2[100]. Another variant is the adaptation of MAPPIT to cells which do not express high levels of STAT3. Hematopoietic cells are a representative example where STAT3 is expressed at very low levels while STAT5 is highly expressed. In this case, instead of being coupled with gp130, the prey plasmid includes a fragment of the βc (beta common) receptor which bears STAT5 recruitment sites[101].

As mentioned above, the MAPPIT technology can detect tyrosine phosphorylation-dependent interactions. To enable the detection of other modification-dependent interactions, Heteromeric MAPPIT has been introduced[102] (**Figure 26b**). Here, the chimeric receptor bait contains, instead of the EpoR extracellular domain, the extracellular domains of each of the subunits of the heteromeric granulocyte/macrophage colony stimulating factor receptor (GM-CSFR): the α chain with the ligand binding site and a β_c chain responsible for signal transduction. While one ectodomain is coupled to the bait protein, the other one is fused to a modifying enzyme: after expression, the bait is enzymatically modified and allows the modification-dependent interaction with the prey. Ligand stimulation remains responsible for the STAT3 activation, detectable in the MAPPIT tests. This method was first used by Lemmens *et al.* to detect serine phosphorylation-dependent interactions in TGF-βR (transforming growth factor-β receptor) family signalling[102].

Another variation of the MAPPIT system is the so-called Reverse MAPPIT (**Figure 26c**). In this case, a positive read out is generated upon interference with a PPI, i.e. the dissociation of a protein moiety able to inhibit the active form of the cytokine receptor complex. Here, the leptin receptor is functional but the prey is bound to a phosphatase which inhibits JAK/STAT signaling. Competing proteins or small organic proteins can destroy the interaction between the inhibitory prey and the bait, leading to the reactivation of the signaling pathway.

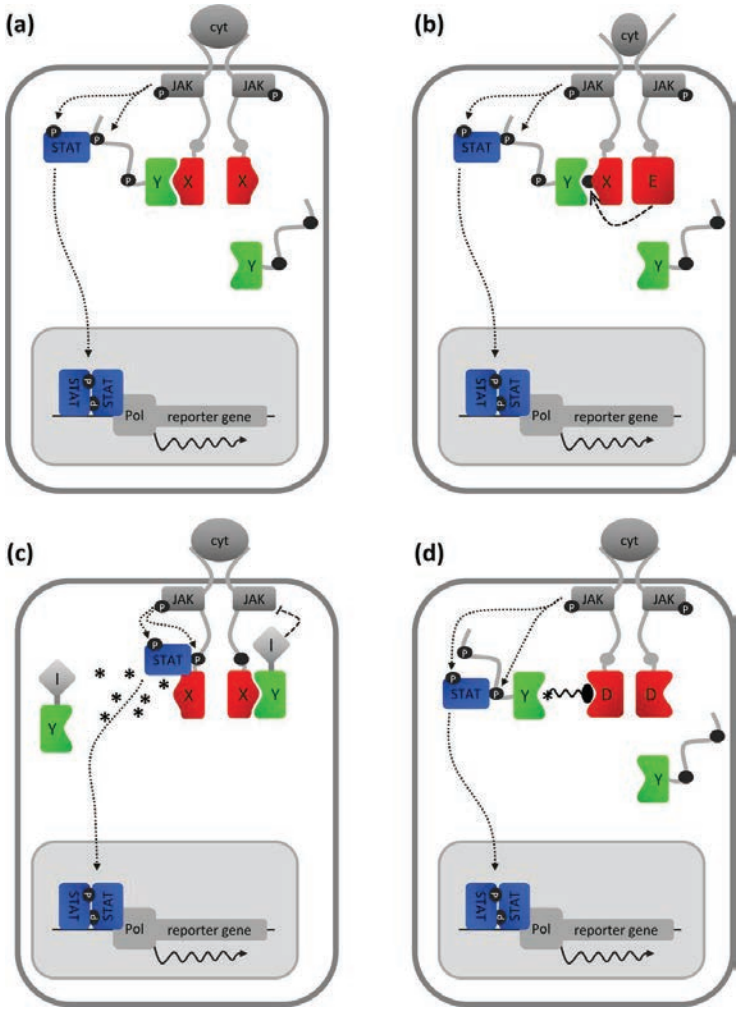


Figure 26 Overview of the different MAPPIT approaches.

(a) MAPPIT: the bait X is anchored to a mutated cytokine receptor containing a mutated STAT recruitment site unable to mediate cytokine signalling. The prey Y is fused to a fragment of gp130 showing functional STAT docking sites (black dots). If X and Y interact, the signaling is reconstituted and STAT3 undergoes tyrosine phosphorylation. Phosphorylated STAT3 migrates to the nucleus, binds to the rPAP-1 promoter and activate the transcription of the reporter gene. (b) Heteromeric MAPPIT: the system uses the extracellular domain of GM-CSFR composed of two chains. The bait is fused to one chain while the other chain is coupled with a modifying enzyme (E) able to modify the bait. After such modification, the prey can recognize the bait, interacts and activate the JAK/STAT signaling cascade.

(c) Reverse MAPPIT: in this case the receptor is functional but the prey Y is linked to a domain (''I'') which inhibits the JAK/STAT signaling cascade. Using a competitor or a small molecule the interaction between bait and prey can be disturbed and the signaling reactivated. (d) MASPIT: it is a three hybrid approach where the chimeric receptor is coupled with DHFR ("D"). Methotrexate (MTX) fusion compound contains a small molecule of interest coupled with MTX. MTX binds to DHFR presenting the small protein as bait. The interaction with a prey protein Y results in the transcription of the reporter gene[103].

Besides all these variations, there is also a three-hybrid system, named MASPIT[104] (**Figure 26D**), which allows organic molecules to be displayed as bait through the use of chemical dimerizers. The chimeric receptor is fused to DHFR (dihydrofolate reductase) : it traps a methotrexate (MTX) fusion compound (MFC) in which a small molecule is connected to MTX via a polyethylene glycol (PEG) repeat linker. This fusion compound thus includes the organic molecule able to act as bait. If the binding of a protein prey to the fusion compound occurs, then the signaling cascade of STAT3 can be initiated. MASPIT also allows screening for novel protein targets of a certain organic molecule. For the analysis, both the FACS-based cDNA library screening or the array-based procedure can be used (see below).

The applications of the MAPPIT technology cover a wide range of possibilities.

Among the applications of MAPPIT, the technology has been performed for pathway walking analysis. Ulrichts *et al.*[105] performed binary MAPPIT experiments to investigate the Toll Like Receptor (TLR) pathway: TLRs are a family of proteins which play a role in the innate immunity. The signaling pathway is activated by the interaction of the receptor complex with one or more adaptor proteins which are recruited via homotypic TIR– TIR (Toll/IL-1 receptor domain) domain interactions. MyD88 (Myeloid Differentiation primary response protein 88), Mal/TIRAP (MyD88-adaptor-like protein/TIR domain containing adaptor-like protein), Trif/Ticam-1 (TIRdomain containing adaptor inducing interferon-b/TIR containing adaptor molecule), Tram/Ticam-2 (Trif-related adaptor molecule/TIR containing adaptor molecule-2) and Sarm (sterile alpha and HEAT-Armadillo motifs containing protein) are examples of TIR-containing adaptor proteins. In this study, MAPPIT technology was used to investigate TLR adaptor recruitment: all known TLR adaptors were screened as baits or preys along with the intracellular part of TLR4 as bait to detect interactions downstream (the homodimerization of the receptor with IRAK4[106]).

The MAPPIT technology can also be applied for drug screening. As a proof-of-concept study, Reverse MAPPIT has been used to detect the disruption of the interaction between MDM2 and p53 by Nutlins [107]. The E3 ubiquitin-protein ligase MDM2 interacts with p53 leading to ubiquitination and protein degradation in the proteasome of p53, abolishing the apoptotic functions of p53 and the cell-cycle arrest[108]. This interaction became subject of therapeutic studies since its inhibition in cancer cell lines leads to induction of CDKN1A expression, cell cycle arrest and apoptosis. Nutlins are cis-imidazoline analogues able to inhibit the interaction between p53 and MDM2 as they bind a hydrophobic region of MDM2 responsible for the p53 interaction[109]. Technically, the receptor is

capable of STAT activation while the prey includes a module which interferes with the JAK-STAT signaling: interaction of the inhibitory prey and the bait results in inhibition of reporter induction. Competing proteins and small molecules can be used to disrupt the interaction: this will restore the signaling, thus leading to a positive readout. Different interactions have been monitored with this technique including EpoR and SOCS interaction or FKBP12 and ALK-4[107].

Drug profiling is another field where MAPPIT technology can be successfully applied. MASPIT (MAmmalian Small molecule Protein Interaction Trap), for instance, was used to evaluate the ABL kinase inhibitor PD173955[110]. In this study, a cDNA library screen was performed using PD173955 as bait and several additional tyrosine kinases were identified. Finally, MAPPIT technology can be applied in large-scale interactome mapping.

To summarize, a key advantage of the MAPPIT technology is its ability to study the molecular interactions in the optimal physiological context for human (or mouse) protein interactions since it is performed in mammalian cells. The commonly used cell type is Hek293T, but different cell types, including epithelial, hematopoietic and neuronal cells can also be used. Dynamic, weak and transient interactions can be detected by the system and the procedure is fast, makes use of simple read-outs (making it suitable for large-scale analyses) and can be largely automated (for Array-MAPPIT), which gives a convenient cost-benefit.

Recently, a new *in situ* technology was developed on the basis of the MAPPIT system concept. A limitation in the original MAPPIT is the fact that the bait is linked to the plasma membrane: this means that some interactions cannot occur in their native localization and full-size transmembrane proteins cannot be used as baits. Furthermore, MAPPIT applications revealed an intrinsic artefact: TYK2-containing prey fusion proteins were identified as constitutive positive signals from multiple screening campaigns. To overcome those limitations a variation of the MAPPIT technology called KISS or KInase Substrate Sensor has been developed (**Figure 27**).

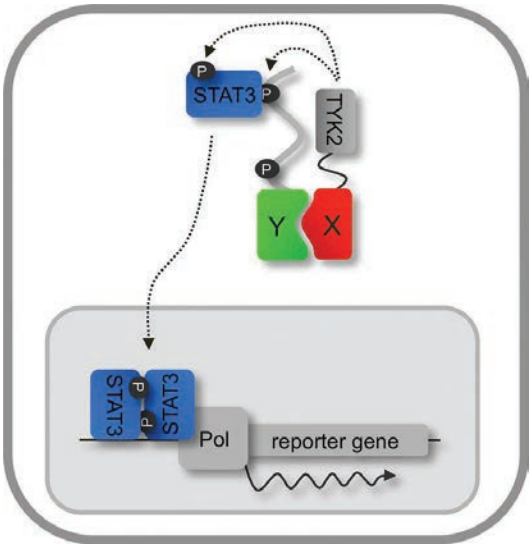


Figure 27 KISS.
In the KISS method the bait is coupled to a constitutively active tyrosine kinase able to activate STAT protein upon interaction of X and Y. The activation of STAT3 leads to reporter gene expression. Picture from [111].

The fact that preys containing TYK2 fused to gp130 show a cytokine-independent MAPPIT reporter activity, means that TYK2 can phosphorylate both the STAT3 binding sites in the gp130 fragment and STAT3 itself. TYK2 has this inherent ability provided it is fused with gp130 in *cis*. By designing the assay in such a way that bait-fused TYK2 and prey-linked gp130 are provided in *trans*, the luciferase reporter signal depends on bait-prey interaction. So in this new method, the prey is the same as this in the original MAPPIT assay, but the bait is coupled with a constitutively active tyrosine kinase: if bait and prey interact, STAT3 phosphorylation occurs followed by the activation of the luciferase reporter[111].

Different studies have been performed to show the broad range of interactions detectable by KISS, involving proteins with different topological characteristics and occurring in different subcellular compartments. For example, cytoplasmic interactions between p51 and p66 subunits of HIV-1 RT were detected with KISS. PPIs involving G protein–coupled receptors (GPCRs) were demonstrated using the interaction of somatostatin receptor 2 (SSTR2) or angiotensin receptor 1 (AGTR1) with beta arrestin 2 (ARRB2) implying correct localization at the membrane of GPCR-TYK2. Interactions between integral transmembrane proteins can be detected as well with KISS: the technique was successfully used to show the ER stress-induced oligomerization of the transmembrane ER stress sensor ERN1[112].

II.2.3.2 Biochemical methods

II.2.3.2.1 Coimmunoprecipitation (CoIP)

The term immunoprecipitation (IP) refers to a technique used for antigen detection and purification using a specific antibody. Detection can then be done either by western blot analysis or by mass spectrometry. CoIP follows the same basic idea and is used also to purify molecules bound to the antigen: in this way, intact protein complexes can be immunoprecipitated (**Figure 28**).

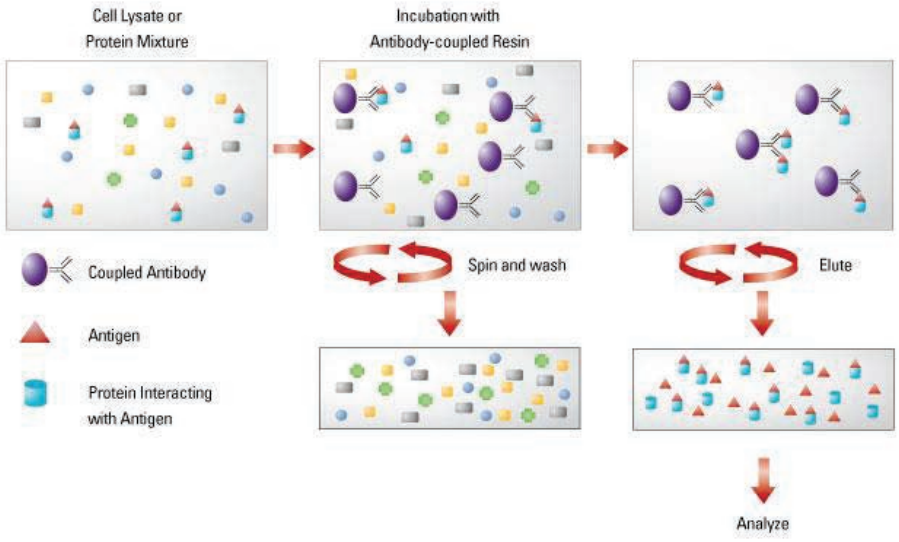


Figure 28 Summary of a standard coimmunoprecipitation assay.
Picture from https://www.thermofisher.com

CoIP is a commonly used biochemical technique for the analysis of PPIs, including interactions of subunits in a protein complex. The analysis is most often performed on a whole-cell extract and the use of eukaryotic cells allows posttranscriptional modifications, which may be important for formation or maintenance of the PPI. To immuno-purify the target protein together with all its interacting partners, an antibody (monoclonal or polyclonal) directed against the antigen of interest is added to the cell lysate and will form an immune complex with the target. This complex can then be purified after being captured by an immobilizing support such as protein-G or protein-A Sepharose: those proteins can bind the IgG from a variety of mammalian species through their Fc

region[113]. The eluate is analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and this is often followed by a western blot analysis to confirm the identity of the antigen.

Obviously the collection of available antibodies doesn't cover the complete list of antigens that can be considered as protein of interest. To circumvent this limitation, a target protein is often fused to a protein tag like glutathione S-transferase (GST), or to an epitope tag such as HA, HIS, V5. When using a GST-tagged protein the assay is called a GST pull-down assay. This assay is used when bait-protein specific antibodies are not available or when it is difficult or impossible to find a cell line able to express the bait protein. An affinity system is then used instead of an antibody: i.e. a glutathione resin able to bind the GST fused proteins. It is important to underline that this is a heterologous system, which means that it might show some downsides such as: the protein may be problematic to express, the protein cannot be purified because it is not soluble, and the protein might be non-functional when expressed. On the other hand, using a tag gives the advantage that it can be easily fused to different proteins so that the same antibody can be used economically.

The fact that the concept of a CoIP experiment is straightforward does not mean that all CoIP experiments need to follow a single protocol. More specifically, relevant aspects need to be considered including the nature of the interaction, the nonspecific binding to parts of the IP assay and the antibody-dependent background. A critical step of the CoIP assay is the stability of the interactions during the washing steps since mechanical and chemical stresses are applied on the complexes during their enrichment. In the case of labile complexes, the composition of the lysis buffer needs to be optimized. To avoid disruption of the interactions, sonication and vortexing are non-advised. A solution for unstable complexes is to perform a crosslinking step: with this approach proteins belonging to the same complex are covalently cross-linked and the bait protein can be precipitated without the risk of losing interaction partners during the CoIP. A common problem of the CoIP assay is the interference of the antibody bands during the gel analysis: the light chain and the heavy chain of the antibody are co-eluted with the targets and show a 25- and a 50-kDa band in reducing SDS polyacrylamide gels, which could overlap with the protein bands under study. To avoid the presence of the antibody chains in the eluate, the antibody can be immobilized to the support, allowing a more specific elution of the target and its binding partners[114,115].

To validate the results of CoIP experiments different controls are needed. Use of a monoclonal antibody will avoid the risk of non-specific pulldown by a heterogeneous mixture of antibodies, but if a polyclonal antibody is unavoidably the choice, it can be used first in a comparable lysate that does not contain the target protein, in order to remove reactive but non-specific antibodies. It is also

important to prove that the co-precipitation requires the antigen of interest; for this purpose a control sample without lysate can be used in parallel. A pulldown with an irrelevant antibody of the same species and isotype as the specific antibody, is a good negative control to test the system.

In the classic CoIP protocol, Sepharose beads are used as solid support: the main characteristic of this resin is the high porosity, perfect for the original aim of purifying milligrams of proteins in columns, but much less for small-scale techniques such as CoIP. To avoid antibody background and loss of proteins during the washing, magnetic beads have been developed. Those beads are small and nonporous to avert the occurrence of hidden faces where the antibody could bind aspecifically; so less washing steps will be required. The properties of the magnetic beads make the experiments more productive and reproducible. The disadvantage of these magnetic beads is the cost, which is considerably higher when compared to the sepharose beads.

II.2.3.2.2 Luciferase based CoIP methods

LUMIER is a luminescence-based mammalian assay that combines a genetic and a biochemical method. Co-affinity purification is used to enrich a protein from the extract and consequently its interactors. This principle was used and optimized to obtain an assay based on a luciferase read out.

A protein is fused to a FLAG-tag and the other contains the *Renilla* luciferase enzyme. An anti-flag antibody is used to precipitate the protein complex of interest and a luciferase read-out is used to show the presence of an interaction (**Figure 29**). Barrios-Rodiles and colleagues developed the technique to systematically map PPI networks in mammalian cells and they applied it to the transforming growth factor–β (TGFβ) pathway[116]. A disadvantage of the method is that FLAG-tagged proteins cannot be quantified. Recently a new version of the technique, called LUMIER with bait control (LUMIER with BACON) was developed. The FLAG-tagged proteins are systematically quantified by ELISA[117,118].

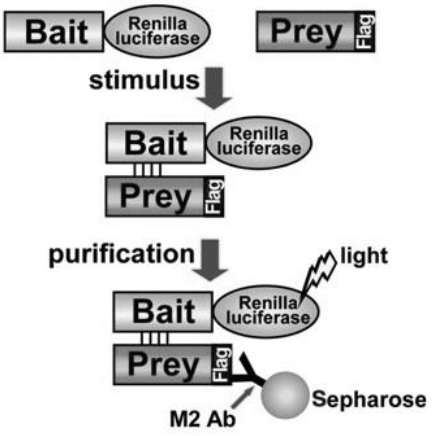


Figure 29 LUMIER is luminescence-based strategy for the detection of mammalian protein-protein interactions
Luciferase-tagged bait is coexpressed with a Flag-tagged prey. Detection is detected as light emission. Picture from [116]

Another dual luciferase reporter pulldown assay used for PPIs is DULIP, for dual luminescence-based co-immunoprecipitation assay. This method allows quantification of both bait and prey hybrid proteins using two different luciferase tags. The bait protein includes a protein A tag for co-precipitation of the complex in the microtiter plates. Quantitative interaction maps can be calculated through normalization of the interaction ratios for all the tested proteins. In 2015, Trepte and colleagues used this technique to investigate the effect of point mutations in the interaction strength of synaptic proteins[119,120].

II.2.3.2.3 Affinity Chromatography

Affinity chromatography separates proteins on the basis of the reversible formation of protein-ligand complexes. This technique shows high resolution and high selectivity and usually results in 90-99% purity. A target protein is immobilized on a solid support in a column and the cell lysate is added: only specifically binding proteins will be retained and in principle all the remaining proteins can be simply washed away. The bound proteins can be eluted by high salt concentration solutions, chaotropic solvents or SDS [115,121]. It is convenient to label the extract *in vivo* with tags. This step will favor detection with high sensitivity and the non-labelled binding proteins can be excluded. This technique was used for the first time in 1976 to detect phage and host proteins that interacted with different forms of *E. coli* RNA polymerase[122]. The purification power of the experiment can be enhanced by the use of a second affinity tag. In case of affinity chromatography high-throughput screening, a particularly interesting multi-tag is the tandem affinity purification (TAP) tag[123]. The TAP- tag is a composite tag, which includes two different epitope domains and a protease cleavage site: it is composed of a calmodulin binding peptide (CBP) followed by the Tobacco etch virus (TEV) protease cleavage site and two IgG binding units of protein-A of *Staphylococcus aureus.* The strategy is based on a dual-purification process: a TAP-tagged protein is expressed in-cell and forms complexes with the endogenous proteins; the purification occurs via a two-step affinity procedure. During the first one, the TAP-tag fused proteins are purified from the lysate by selection on an IgG matrix, retaining all the protein-A-tagged proteins. The TEV protease can then cut the bound tag. In the second step the protein complex is immobilized to calmodulin-coated beads via the CBP of the TAP-tag. The result of the process is a highly purified extract of proteins with very low background. Moreover, the non-denaturating conditions of the assay enhance the possibility to obtain intact protein complexes[121,123–125]. Next, the purified protein sample undergoes separation by an one-dimensional SDS/PAGE and analyzed in MS (**Figure 30**)[126,127]. Mass-spectrometric analysis allows

determining the molecular masses through the measurement of the mass to change (m/z) ratio of charged molecules (ions) in gas phase.

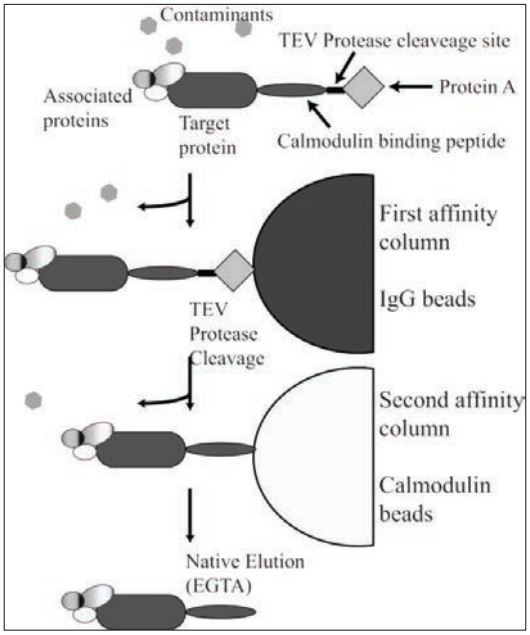


Figure 30 Schematic representation of the TAP–MS methodology.
A construct with double tag is used, including a calmodulin-binding peptide, TEV protease cleavage site, and two IgG-binding units of protein A. The protein A is used as tag to recover the fusion protein from the extract, a protease cleavage allowed the release of bound IgG material. Extra purification step is performed to remove contaminants using the affinity of calmodulin-binding peptide for calmodulin beads. Ethylene glycol tetra-acetic acid EGTA released with the proteins offer high-purity of the recombinant proteins. (Figure from [128]*)*

Tags

Proteins produced in a different host could undergo a degradation process. On the other hand, cells that undergo a transfection/transduction process are forced to overproduce a large amount of protein, which can activate the proteasomal pathway. Introducing a tag has positive and negative sides. Affinity tags, besides preventing proteolysis and increasing solubility, may improve protein yield and facilitate protein refolding. They have also been used to increase the sensitivity of binding assays for tagged single-chain fragments of antibody variable regions (ScFv)[129]. On the other hand, it has been shown that affinity tags can interfere with the quality of the protein produced, for example, by changing the biological activity, by inhibition of enzyme activity, by alteration of the protein conformation or by causing toxicity[128]. Specific cloning vectors are used to create fusion proteins considering the right reading frame. Glutathione S-transferase (GST), for instance can be used as tag in an one-step purification experiment with glutathione (able to protect proteins from proteolysis). Other tags have been used to enhance stability and solubility of the proteins, e.g. maltose-binding protein (MBP), thioredoxin A (TrxA), small ubiquitin related modifier (SUMO),

ketosteroid isomerase (KSI), and TrpΔLE. The use of a tag becomes crucial in case of uncharacterized proteins or in the absence of a good antibody for the protein of interest, as is often observed in systematic, large-scale interactome studies. Flag, HA and c-myc are the most commonly used tags for western blot analysis, IP, CoIP, immunofluorescence (IF), and flow-cytometry. The earliest affinity tag developed was an epitope of the human c-myc proto – oncogene product (EQKLISEEDL), interacting with a specific c-myc 9E10 antibody[130]: it has been used to purify c-myc fusion proteins for structural studies[131]. The HA-tag is a nine amino acid sequence (YPYDVPDYA) derived from the human influenza virus hemagglutinin protein. Despite the fact that the epitope might be spatially located, its introduction can interfere with folding, trafficking or normal functions of the protein. The FLAG epitope is a short, hydrophilic octapeptide (DYKDDDDK), including an enterokinase-cleavage site, that can be used for antibody-based purification and it was the only epitope tag to be patented[132]. An added advantage of this tag is its high hydrophilicity with a maximum score on the hydrophilicity scale according to Hopp and Woods.

II.2.3.2.4 Proximity-ligation assay (PLA)

PCR-based technologies, used for the detection of nucleic acids, and antibody-based methods applied for protein studies have been two key technologies for the advances that life science research has made over the years. Original combined applications of both techniques lead to the development of immuno-PCR (iPCR) and, proximity ligation (PLA) and extension (PEA) assays, which use the specificity of detection given by an antibody, with the capacity of amplification given by PCR or RCA (see below).

The PLA assay (**Figure 31**) uses DNA oligomers and DNA modifying processes such as ligation and cleavage. In the original PLA method, the protein of interest is tagged with a pair of proximity probes. A proximity probe pair is composed of two oligonucleotide-labelled antibodies able to bind in close proximity (30-40 nm apart) to different epitopes of a protein or two proteins in a complex. This system can follow a direct or an indirect protocol: in the first case, detection is performed by direct primary antibody while in the second, secondary conjugates. After binding to a target molecule or interacting molecules, the proximity antibody-conjugated oligonucleotides can hybridize to two additionally introduced connector oligonucleotides to form a ligation product. Amplification and detection can then be done in two ways: via a qPCR detection which occurs simultaneously to the amplification and which requires linear amplification products, or using rolling circle amplification (RCA) if the ligation creates a circular single stranded DNA molecule.

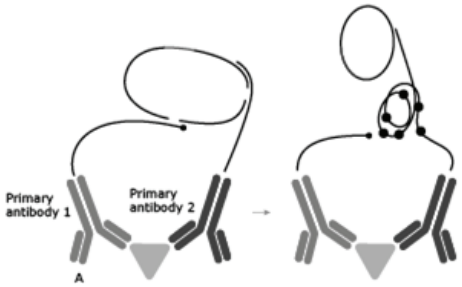


Figure 31 Proximity ligation assay.
Two probes stay in close proximity by binding to a protein or two proteins present in one complex: upon introduction of linear connector oligonucleotides, they can be joined and circularized. The rolling circle amplification can start and one of the probes is used ad primer. The products are hybridized with labeled detection oligonucleotides
Picture from Human protein atlas[133].

RCA is an isothermal enzymatic reaction which is based on particular DNA polymerases such as Bst and Vent exo-DNA polymerase for DNA, and T7 RNA polymerase for RNA, to obtain a long repetitive single-stranded amplicon showing up to 1000 of reverse complementary elements of the circular template. The product of the reaction is detected and visualized via hybridization of complementary fluorescence-labelled oligonucleotides[134–136]. The method has been used to show the known interaction between endogenous c-Myc and Max: the interaction was visualized at a single molecule level[134]. PLA experiments can be carried out with samples and reagents in a single tube as a homogenous assay thus avoiding variability due to successive washing steps. It is also possible to conduct the experiments in a solid phase if proteins are immobilized in a solid support such as affinity beads and it is mainly used for direct detection of proteins in biofluids. The *in situ* PLA is the method to study PPIs described from Soderberg *et al*. for the first time in 2006 [134] and in this case the assay is performed in fixed cultured cells or tissue sections. Due to the fixation, transient interactions can be detected but real time analysis of the interactions in living cells is not possible.

PLA has been used as basis to develop a high-throughput biomarker screening technology. Numerous PLA probes can be bound to different combinations of oligonucleotides resulting in a sequence that can be used as primer for qPCR amplification. This multiplex PLA has been extended following the principle of homogenous multiplex proximity ligation assay up to four 24-plex panels which contains 74 biomarkers. The sensitivity of the test is very high and after optimization of the technique, including pre-amplification strategies, very low sample quantities are required. The assay has been validated in plasma samples of 74 patients with colorectal cancer and in a sample set of 74 healthy individuals[137].

II.2.3.2.5 Other biochemical methods

A few more biochemical methods can be applied to identify or characterize PPIs. Here, the two most commonly used biochemical methods will be briefly described: protein microarrays and cross-linking approaches.

Protein microarrays or Protein chips can be performed for a large variety of applications. Initially, this approach was applied for screening of protein biochemical activities[138], however further applications of the approach include the identification of PPIs, protein-phospholipid interactions, the small molecule target identification, and the search for substrates of protein kinases. Furthermore, protein microarrays can be used in the clinic for diagnostic applications and for monitoring of disease stages[139]. The classic approach is the so-called analytical protein microarrays, which are mostly used to analyze a mixture of proteins for binding affinities, specificities and expression levels of its components. A glass microscope slide is used as the solid surface and a library of antibodies is arrayed on the slide. The array is then probed with a heterogeneous protein solution[140]. There are two types of protein microarrays related to the analytical one: the functional and the reverse-phase protein array. In the first case, the array includes full-length functional proteins or domains, and it is used to study the activities of an entire proteome. This approach has been used to study various interactions such as protein-protein, protein-DNA, protein-RNA, protein-phospholipid, and protein-small molecule interactions[138,139]. In the reverse-phase protein array, the lysate is on the slide and specific antibodies are added on it. The antibodies are typically detected with chemiluminescent, fluorescent, or colorimetric assays. This variation enables the determination of disease-related altered protein mixtures. For example, reverse-phase protein microarrays can detect post-transcriptional modifications, which may occur in pathological conditions[141].

As already mentioned, an advantage of the protein array technology is that it can be used with many types of molecules as probe enabling the study of protein-DNA or protein-small molecule interactions, beside PPIs. The downside of the protein chip is that it is an *in vitro* approach and that it shows some technical limitations such as, for example, the difficulty to purify the numerous proteins under native conditions and to spot them accurately on a surface.

Cross-linking techniques have been used to allow the study of transient and low affinity interactions. Proteins able to approach each other in a close range can be covalently cross-linked in living cells allowing the study of their interactions. There are different kinds of chemical cross-linking agents that can be used for this purpose; the most commonly used are the imidoester cross-linker dimethyl suberimidate, the N-Hydroxysuccinimide-ester cross-linker BS3 and formaldehyde. Formaldehyde is

most frequently used and is a reactive dipolar compound that cross-links proteins via the amino- and imino-groups of their lysine, arginine and histidine residues. It is a fast and perfectly reversible reaction, which allows an affinity purification or immunoprecipitation step before MS identification [142–144]. In 2005, Suchanek and colleagues introduced a new cross-linking strategy based on photo reactive amino acid analogs: two photoactivatable amino acids, such as photo-methionine and photo-leucine, are incorporated into the proteins and activated upon UV light exposure, which induces covalent cross-links. Detection of protein complexes is done via western blot[145].

II.2.3.2.6 BioID

Most of the biochemical methods to detect PPIs, such as CoIP and TAP-Tag, include steps (such as detergent-mediated cell lysis and subsequent washing steps) which can compromise the stability of weak interactions or cause the loss of the transient ones. Some strategies have been developed over the years to overcome this problem, for example the use of the cross linking with a down-side of creating large insoluble complexes. Recently, different studies focused on the possibility of enzyme-catalyzed proximity labelling to be used for PPI detection. Some enzymes can be used as a label for the protein of interest and their short life allows an activity limited to proximity. Different enzymes have been used for this aim[146]. BioID (proximity-dependent biotin identification) was originally developed by Roux and colleagues to study the surrounding neighbors of a protein of interest based on a promiscuous prokaryotic biotin protein ligase[147,148]. The *E. coli* enzyme BirA (BirA R118G) is a 35-kD DNA binding biotin protein ligase able to regulate the biotinylation of a subunit of acetyl-CoA carboxylase. BirA shows an extreme specificity in biotinylating its substrate peptide requiring a specific recognition sequence; it combines ATP and biotin to produce biotinoyl-5'-AMP, a highly reactive intermediate with very short life. The method uses a modified enzyme, BirA* which can biotinylate any protein with an exposed lysine residues, found in proximity of the enzyme (about 10 nm) missing the specificity of BirA. The mutated enzyme is first fused to the protein of interest and then introduced in the cells. In presence of medium with biotin, the BirA* will biotinylate proximal proteins which can be purified by streptavidin binding and identified via MS[149]. However, this method does not guarantee the identification of the physical interactions between the protein of interest and the candidates, but rather the proximity to the bait (**Figure 32**).

Initially, BioID was applied to investigate the neighbor proteins of Lamin A, an intermediate filament protein, which play a role in the cytoskeletal of the nuclear lamina. The fusion protein was introduced in Hek293 cells. 100 proteins (including well known lamina-binding proteins and new interaction partner) were biotinylated after sub-ministration of biotin[147]. Recently a new improved

smaller biotin ligase has been proposed and used to improve the ability of labelling proximate proteins[150].

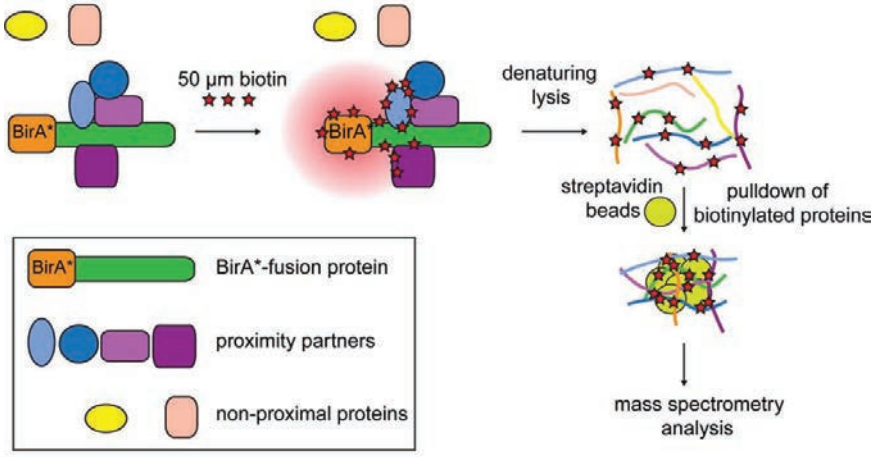


Figure 32 BioID: the principle
The protein of interest is tagged with a modified promiscuous biotin ligase BirA (BirA) and introduced in the cells. An incubation of about 24 hours with medium supplemented with biotin allows the biotinylation of the proximity interactors. The radius of action is about 10 nm. Picture from* [149]

.

II.3 Interactome Mapping

Recently, the exploration of the so-called "protein interactome" became possible thanks to the development of new and innovative technologies based on the analysis of PPIs at a high-throughput level. An interactome is defined as the complete set of interactions within an organism. In the early days two experimental approaches have been mainly used to start mapping the complete sets of PPIs: the yeast two-hybrid screening method and the detection of protein complexes by affinity purification and mass spectrometry. Next to these classic ones, new techniques have been developed including PCA-based methods and more recently also high-throughput applications of MAPPIT. These main methods will be presented in the following paragraphs.

II.3.1 Yeast two-hybrid (Y2H) screening

High-throughput two-hybrid screening utilizes yeast mating. Two haploid strains are involved: one in which a query or bait protein is fused to a DNA binding domain (BD) able to bind to an engineered site upstream a reporter gene, and another one in which the interaction or prey protein is fused to a transcription activator domain (AD). The measurement of reporter activity in the diploid organisms, resulting from mating between the two strains, is the read out of the assay: if the two proteins interact mutually, the AD activates the reporter leading to the transcription of the reporter gene which will be used for the selection.

The concept formed the basis of developing the Y2H system into a high-throughput screening tool during the early 1990s. In 1996, the lab of Fields published a paper about the use of the Y2H system on a genome-wide scale to identify interactions between the proteins of the *E. coli* bacteriophage T7[151]. However, at that time, the two-hybrid methodology was not suitable for a large-scale project like the human interactome without specific methodological improvements towards completeness and selectivity of the screens. Large scale interactomic studies have since then been pursued for *Saccharomyces cerevisiae* and for *Arabidopsis thaliana* with the aim of applying the strategy developed also to other genomes and the eventual elucidation of the human protein interactions map[152,153]. Subsequently, more extensive interactome mapping projects have been carried out for eukaryotes like worm and fly[38,39,154,155]. The very first original large-scale analyses aiming at whole proteome coverage were performed in yeast[38,39]. Uetz' studies led to the identification of 957 interactions among 1004 *S. cerevisiae* proteins while Ito's screening identified 4549 interactions among 3278 yeast proteins. It is important here to notice that there was a limited overlap between the data from the two experiments. Successive studies showed how Y2H screens generally generate

high quality binary data, indicating that the observed low level of overlap between various sets of results is not due to the low quality of positive hits but due to assay sensitivity and screening completeness thus precision needs to be considered[156] (more details in paragraph II.3.6 of this dissertation). With time, it became clear that the identification of the complete collection of all physical interactions that take place within a cell will require the creation of genome-scale resource collections of open reading frames, the so called ORFeome. An ORF is the entire coding sequence between the initiation and termination codons, excluding the untranslated 5'UTR and 3'UTR of mRNAs[157]. Ideally, any cloned ORF should include the different isoforms, and gene's variants. Technically, the exhaustiveness of an ORFeome resource mainly depends on the quality of the genome annotations. Importantly, the ORF collection can be easily transferred in the plasmid vectors required to apply the different techniques. The first human ORFeome version 1.1 dates to 2004[158]. In this work, researchers cloned all unique human ORFs that were at that time available as full-length cDNAs in the Mammalian Gene Collection[159]. The human ORFeome v1.1 collection[158] has been used to perform the second large scale Y2H study for humans: 7200 different protein-coding genes were investigated and 2754 interactions were detected. Co-affinity confirmation tests yielded positive results for 78% of the candidate PPI.

As already discussed, yeast remains a very good model to study human biology: despite the fact that the last common ancestor dates back a billion years ago, humans and yeast still share a substantial part of their genetic heritage. The yeast genome includes about 6000 protein coding genes, less than a third of the human amount, but 2300 of those are human orthologs. Most of the common genes show important roles for the cells in both organisms and mutations or perturbations of them is the cause of a range of human disorders including many Mendelian diseases and cancer. This strong similarity is one of the main reasons of the success of the different Y2H techniques and of their development into high-throughput assays. The latest improvement of the technique involves the development of the so-called "humanized Y2H"[160]. There are five different degrees of yeast humanization, shown in **Figure 33** which goes from the use of non-humanized yeast to address questions about human biology to the humanization of the whole yeast genome or humanization of full pathways and complexes.

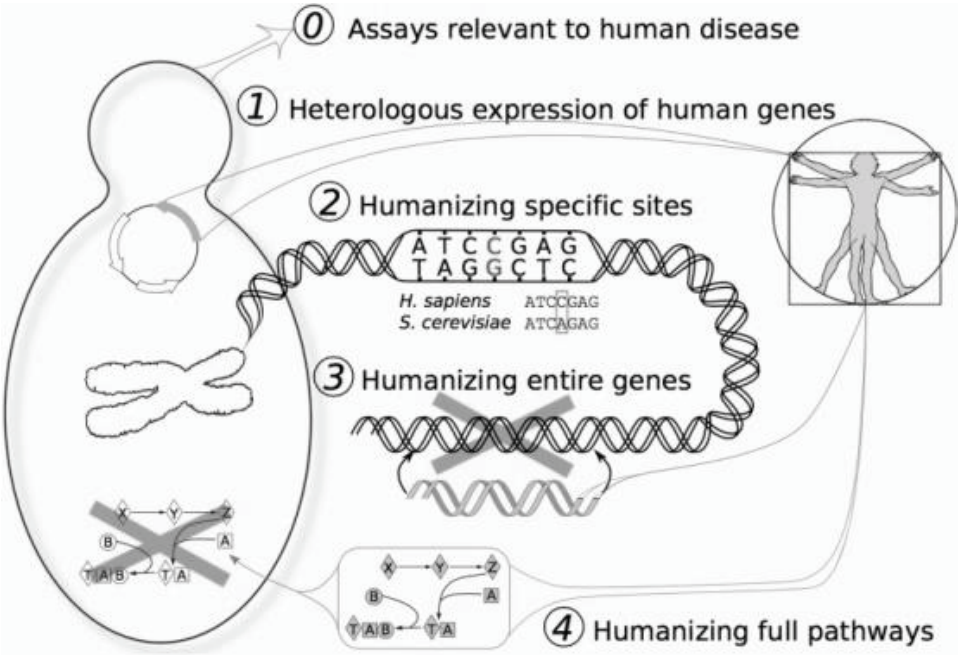


Figure 33 Five degrees of yeast humanization.
Yeast has proven useful for the direct study of human biology in a variety of forms, illustrated here to distinguish those cases in which yeast cells were simply studied for human-specific processes and drugs (degree 0), to the heterologous expression of human genes in yeast (degree 1), all the way to the directed replacement of specific amino acids, genes, and pathways (degrees 2–4, respectively).
Degree 0: Studying non-humanized yeast to address human biology; Degree 1: Expressing human proteins in yeast regardless of orthology; Degree 2: Humanizing specific positions within native yeast genes; Degree 3: Humanizing entire yeast genes; Degree 4: Humanization of full pathways and complexes [161].

The first experiment of humanization dates back to 1985 when Kataoka and colleagues demonstrated the functional homology of mammalian and yeast RAS genes expressing human or chimeric Ras in ΔRas mutant yeast[162]. Since then, over 400 yeast genes have been humanized: the ultimate goal of this system's evolution is the application for therapeutic purposes. Making personalized strains or express any possible allele or combination of genes could be the start to identify critical mutations responsible for a disease of a specific patient and screening of such yeast strains could be performed for therapeutic resistance or response for the specific case[161].

II.3.2 Affinity purification and mass spectrometry

Affinity purification (AP) followed by mass spectrometry (MS) analysis is the second most widely used approach for high-throughput mapping of PPIs. AP is used to isolate those complexes that

comprise a protein of interest. After AP of the tagged-target protein, under conditions that preserve the protein complexes, MS is applied for identification of the proteins.

The first time that a large scale analysis of PPIs by AP combined with MS was reported was in 2002 when two studies on *S. cerevisiae* were published[163,164]. In the first study, Gavin *et al.* combined TAP with MS and could purify 589 specifically interacting proteins; the reproducibility of the purifications was found to be as high as 70%. In the second study, Ho *et al.* purified 725 FLAG-tagged proteins and analyzed the complexes through tandem MS. They obtained data for up to 25% of the yeast proteome. Results were confirmed by immunoprecipitation experiments in 74% of the interactions. More extensive studies on the yeast interactome have meanwhile been performed: these data cover 47% of the yeast proteome but the data are overlapping only for 18% [165,166].

In 2007, the AP approach was applied on the human interactome[167]: 338 protein baits were selected on the basis of known or suspected disease implications and functional associations. A dataset of 6,463 interactions among 2,235 distinct proteins was generated. Since then, different studies were performed to produce a more complete human interactome map. Recently, a new platform has been developed by the Harper and Gygi labs. In the study carried out by Huttlin and colleagues, the AP-MS approach was used to define the BioPlex network: 2,594 baits were investigated offering 23,744 interactions among 7,668 proteins, more than the 10% of the total human amount[168] (**Figure 34**). The most recent update defines an interaction network of 33.000 interactions involving 8.000 proteins and using 3.200 baits. They also developed a new version of the CompPASS method, called CompPASS-Plus, to analyze the data.

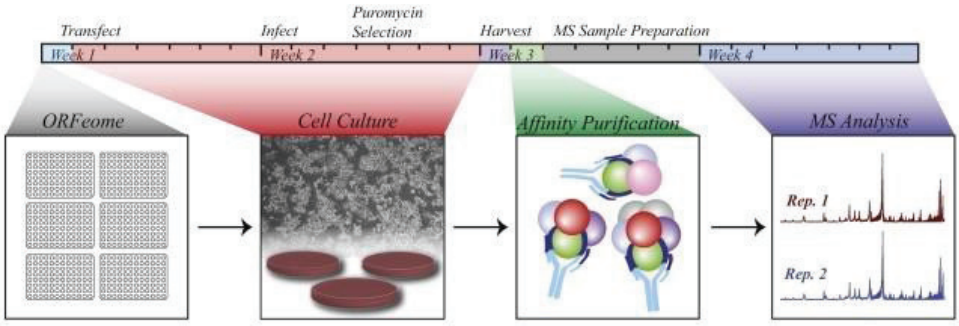


Figure 34 High-Throughput Interaction Mapping via AP-MS
Starting from 13,000 ORFs from the Human ORFeome collection v 8.1 [169], FLAG-HA-tagged baits were prepared to constitute a lentiviral library. Hek293T cells were infected under puromycin selection and the expressed baits were immuno-purified and, analyzed via MS. Modified from[168].

II.3.3 Virotrap and SFINX

One of the limitations of all the affinity based approaches and of the classical mass spectrometry–based strategies is the requirement of cell lysis as an inevitable step. The isolation of complexes from the lysate means that all the compartments of the cell are mixed and some complexes are identified even though they are not likely to happen *in vivo* because of different localization of the proteins. Additionally, the use of tags can increase the number of non-specific complexes[170].

Different strategies have been applied to overcome the risk of false positives caused by background contaminants due to aspecific protein interactions with the solid-phase support, affinity reagent or epitope tags. These include the use of more stringent washing protocols which may lead to the loss of weak interactors, the introduction of the TAP-tag, the creation of a CRAPome (currently available also for *Homo sapiens* and *Saccharomyces cerevisiae* at http://www.crapome.org/) which is a contaminant repository for affinity purification which allows the identification of false positives by comparing the occurrence of such interactions in unrelated AP-MS negative control experiments[171]. To avoid losing interactions, cross linking strategies with simple reagents as formaldehyde or with more advanced isotope-labelled cross-linkers have been applied[163,172]; despite all these optimization strategies, the conditions are still not optimal. In this context and considering the fact that only a small percentage of expected PPIs has been identified with the existing methods[173], a new technique has been developed, the Virotrap [174].

Virotrap is a co-purification strategy able to trap protein complexes under native conditions in virus-like particles (VLPs) that bud from human cells. Due to the evolutionary stress, viruses developed a condensed organization of all information into a small genome. In case of HIV-1, a single protein, the p55-GAG, is sufficient for the production of the virus-like particles[175]. To eliminate the cell lysis step, secreted VLPs are used as traps for protein complexes and as a protective envelop for further purification. The empty VLPs produced in the absence of a virus, contain about 5000 uncleaved Gag proteins each[176]. The bait proteins in a Virotrap experiment are fused to the C-terminus of the Gag protein, which means that the VLP inner membrane is coated by bait fusion proteins and the interacting proteins are trapped inside the VLP after the budding events. Virotrap was first used to demonstrate the known interaction of HRAS and RAF1. The enrichment was performed by ultracentrifugation and then optimized to avoid this step using the co-expression of a tagged version of the vesicular stomatitis virus glycoprotein (VSV-G) that is expressed at the same time as the fusion baits and preys. The HRAS–RAF1 interaction was confirmed using this single-step protocol (**Figure 35**).

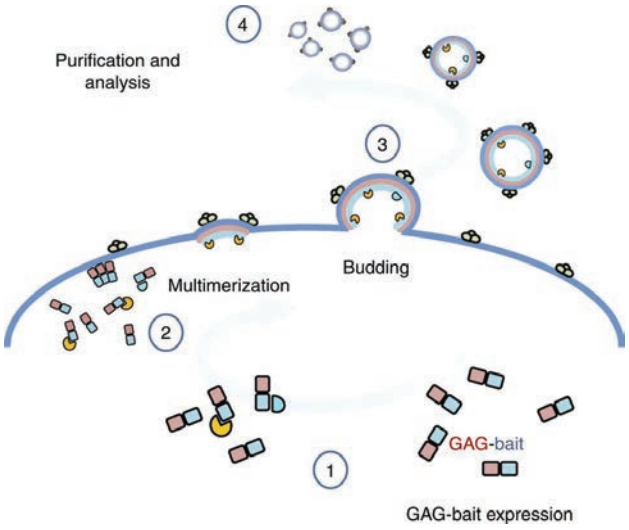


Figure 35 Virotrap principle
Expression of a GAG-bait fusion protein (1) will pick up interactors during trafficking to the membrane. There, multimerization occurs (2) and subsequent budding of VLPs from cells (3) after sufficient Gag-bait fusion proteins are assembled and have reached the membrane. Interaction partners of the bait protein are trapped within these VLPs which are released in the supernatant and purified via a tagged VSV-g protein. Interaction partners are identified after purification by western blotting or MS analysis (4). (Modified from [174])

To compare Virotrap with other technologies, binary analysis of the human positive reference set (hsPRS-v1, 92 PPI pairs) and the corresponding random reference set (hsRRS-v1, 92 randomly selected pairs) described by Braun *et al.* [177] has been performed showing that Virotrap complements other methods and can detect different sets of PPIs when compared to MAPPIT, for example. Every analyzed method is able to detect ~20-35% of known protein interactions. The results only partially overlap suggesting the complementarity of Virotrap with the other methods[174] (**Figure 36**).

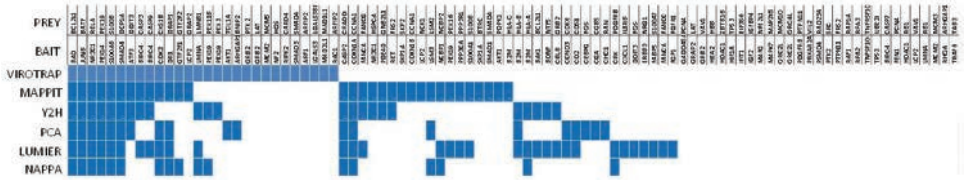


Figure 36 Binary Virotrap assays for the Positive Reference Set (PRS) 1 and comparison against other binary protein-protein interaction assays.
For the purpose HEK293T cells were transfected with PRS and RRS (184 interactor pairs). Single step purification and WB analysis of the eluted particles and lysates of the producer cells were performed. E-tag antibody was used in VLP samples. The blocks show that Virotrap results in 30% positive interactions (at the expense of 5% false positive signals in the RRS set, not shown) (Modified from [174]).

The analyses of the data generated by Virotrap required new analysis tools. SFINX (http://sfinx.ugent.be/), the Straightforward Filtering INdeX for data analysis has recently been developed and can also be used also for AP-MS datasets[178]. SFINX combines the positive feature of different classical methods and offers the identification of true positives in a fast, user-friendly, and highly accurate way. Classical methods such as PP-NSAF, the CompPASS scores, and SAINT, use the spectral count data defined as the total number of peptide-to-spectrum matches assigned to the protein of interest. One of the innovative characteristics of SFINX is the use of peptide counts instead of spectral counts where the peptide count is the number of distinct peptide sequences found for the protein of interest. The protein gets detected as a true interactor if it is sufficiently exceptional for the bait projects and sufficiently reproducible[178].

II.3.4 Protein fragment Complementation Assay (PCA)

PCAs are a family of assays which allow the quantitative probing of molecular interactions in living cells in multicellular organisms, or *in vitro*. The technique has been described in II.2.3.1.4. The assay is based on the PCA principle, where two proteins of interest are fused with complementary fragments of a reporter protein, which is reconstituted in case of PPI. DHFR (dihydrofolate reductase) is an enzyme involved in the biosynthesis process of purines, thymidylate and certain amino acids. Methotrexate (MXT) binds and inhibits DHFR[179]. Two detection methods have been successfully developed: a survival detection assay, using the ability of cells expressing complementary fragments of DHFR to survive in a selective medium and a fluorescent assay based on the detection of fluorescein-conjugated MXT binding to a reconstituted DHFR. Tarassov *et al.*[180] used a DHFR-based complementation assay in a genome-wide *in vivo* assay. In this first screen 1124 yeast proteins were analyzed and 2770 interactions were identified[180] confirming between 16 and 41% of PPIs identified in previous large scale studies.

II.3.5 MAPPIT

MAPPIT is based on the JAK-STAT signal transduction pathway triggered by the interaction between cytokine receptor-derived hybrid proteins[99]. The interaction of bait and prey chimeras leads to the recruitment and successive activation and nuclear translocation of STAT3 dimers. The result is the transcription of a reporter gene in case of the analytical tests or of a selectable marker gene for screening experiments, as these genes are under the control of a STAT3 responsive rPAP1 promoter. Given the simplicity of the assay two high-throughput approaches were developed: screening of

complex prey cDNA libraries using a FACS-based approach, and an array-based assay using the human ORFeome[181].

II.3.5.1 FACS-based approach

For the first approach, a stably-transfected Hek293 cell line expressing a STAT3-responsive rat pancreatitis-associated protein I (rPAP)-puro^{R} selection cassette and a Flp recombination target (FLP) integration cassette in a transcriptionally active locus was generated. In addition, the cells express a murine ecotropic retroviral receptor to enable the retroviral delivery of a prey cDNA library. In this cell line, cell clones expressing the chimeric receptor coupled with the bait can be generated using the FLP integration cassette, via a FLP recombinase assisted integration. This is then followed by an infection with a retroviral prey cDNA library. Upon stimulation with the ligand of the receptor and under the control of the puromycin selection, colony formation is occurring only if bait and prey interact (**Figure 37**).

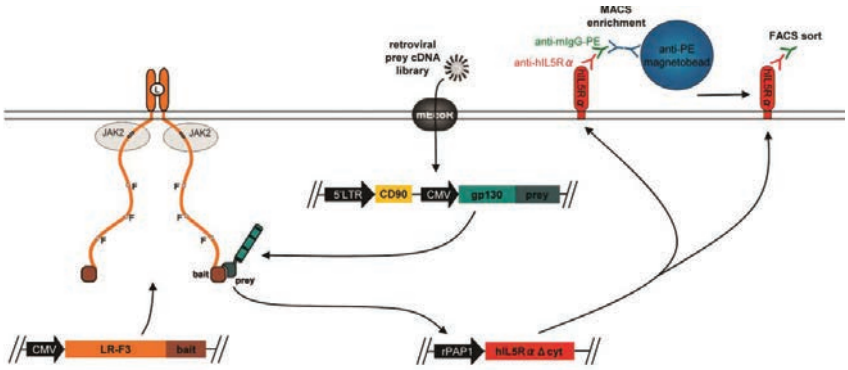


Figure 37 Schematic representation of a MAPPIT cDNA library screening procedure.
A bait-expressing cell pool undergoes a retroviral delivery of the prey cDNA library. After stimulation with the ligand, cells containing bait-prey complex are enriched in a MACS step with magnetobeads attached to the hIL5Ra membrane tag. After a second stimulation, single positive cells are sorted and clones are validated in a dot-blot assay. RT-PCR and sequencing are used to determine the identity of the interacting preys. Figure from[182].

In 2004, Lievens *et al.* optimized the method to avoid the use of puromycin selection, which appeared to limit the screening sensitivity of the technique[182]: in this new method a membrane surface tag is used as a marker, under the control of the STAT3-dependent rPAP1 promoter, enabling selection by flow cytometry. Indeed, instead of using the puromycin cassette, an rPAP1-hIL-5RαΔcyt construct was stably integrated in the cell line. hIL-5RαΔcyt is a fragment of the human interleukin 5

receptor, α subunit, missing the cytoplasmic part[181] to avoid internalization, for which a high-affinity monoclonal antibody is available.

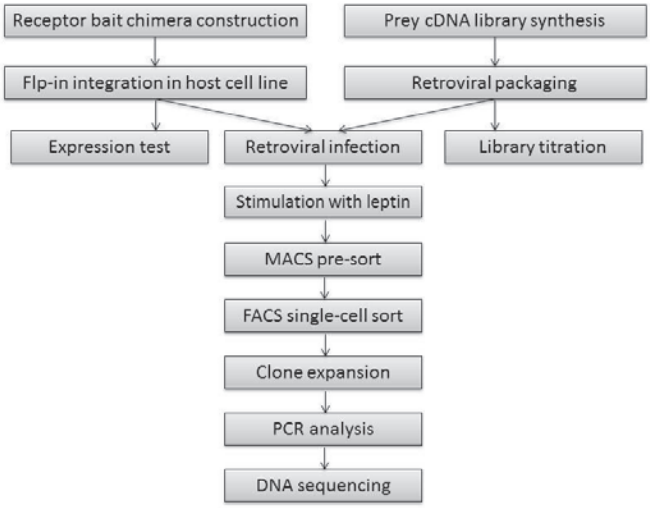


Figure 38 Flow chart of the FACS-based MAPPIT cDNA screen.
See text for details.

Briefly, the bait construct is cloned in a vector for expression of the bait-receptor chimera and transfected into the host cells. After selection of positive clones expressing the bait-chimera, the cell line is infected with a prey cDNA library and stimulated with receptor ligand. A magnetic cell sorting (MACS) enrichment for positive cells is performed before the fluorescence-activated cell sorting (FACS) analysis. Usually a cell sample that has been infected with a mock retroviral library is processed in parallel to be used as a background control. The isolated clones are validated using a dot blot assay, a technique for detecting, analyzing, and identifying proteins, similar to the western blot, where mixtures including the proteins of interest are directly spotted on a membrane without prior electrophoretic separation. The identities of the preys are then determined through RT-PCR and sequencing (**Figure 38**). Due to the large complexity of the library involved, the isolation of positive cells and the identification of preys are difficult and lengthy. To make the process faster and easier, an alternative was developed: the so-called Array MAPPIT, a fast, automated and cost-effective approach.

II.3.5.2 Array MAPPIT

This method is based on the availability of a large collection of full length human ORFs[183], which were transferred in the MAPPIT prey vector to create the library collection. This method requires a protocol based on a reverse transfection: prey and luciferase reporter plasmids are mixed with gelatin and a transfection agent and spotted in microtiter plates or on microscope slides, and dried (**Figure 39**). When needed, (the plates can be stored for a several weeks) the plates are overlaid with bait-expressing cells leading to plasmid uptake and expression of the preys. The stimulation with the receptor ligand activates the bait-receptor chimera, which leads to a correlated and measurable luciferase signal. Unstimulated samples are used as control for the background. The prey identity is simply identified through the position on the plate. The procedure, mostly automated, takes 5 days. The first Array MAPPIT screening was performed in 2009[181] to identify interaction partners of SKP1 and Elongin C. The microtiter plate included 1,879 full length ORF preys selected as they were classified in the 'signal transduction' section in Gene Ontology[184]. The test was performed twice to evaluate the reproducibility and led to the identification of 5 interaction partners of SKP1, of which three known and two unknown, and 5 partners for Elongin C, 4 of which were previously identified. These results were obtained with a very stringent threshold: lowering it implies that the number of detected interactions is increased but also the number of false positives[181].

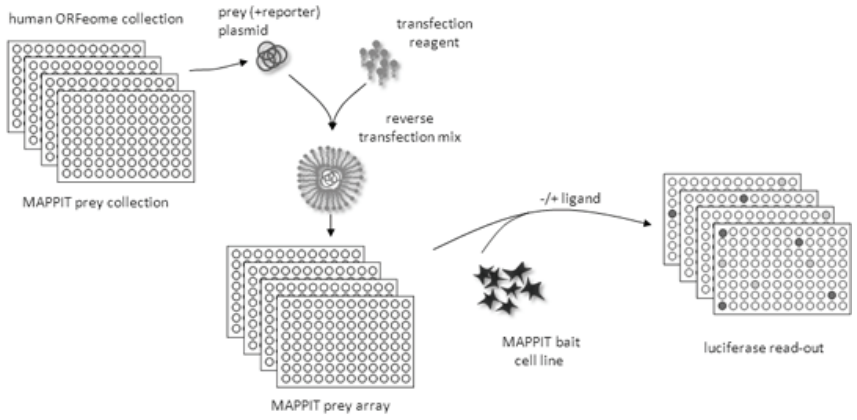


Figure 39 Array MAPPIT.
(See text for details)

II.3.5.3 A MAPPIT/MASPIT cell microarray screening platform

MAPPIT and MASPIT assays, as already presented before in this chapter, operate in mammalian cells and are based on reconstitution of a signaling deficient cytokine receptor. The efficiency of the system has been shown to be similar or better than other well established interaction assays. Recently, a highly miniaturized screening platform[185] for the identification of novel protein interactors of both proteins and small molecules based on MAPPIT/MASPIT platform was developed: the combination of the techniques with a cell microarray format offer an assay suitable for industry application. The screening offer the human ORFeome v8.1 and ORFeome Collaboration prey collections, which show about 15K plasmids and use reverse transfected cell microarrays. The prey collection and the reporter plasmid are printed on a polystyrene plate (**Figure 40**). The system has been validated in a study for RNF41 interaction partners. A mutated RNF41, unable to auto-phosphorylate, was used as bait. The results were ranked according to the normalized fluorescence signal intensity. A number of known interaction partners were identified and a list of novel candidate was proposed: two of them are currently investigated for functional implications.

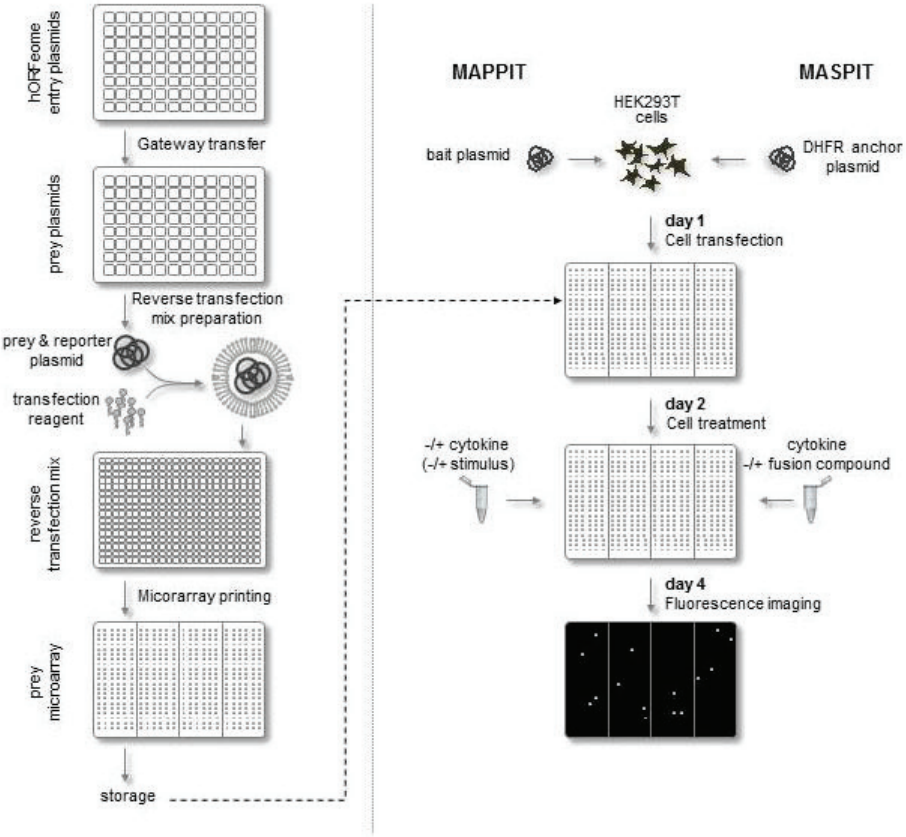


Figure 40 Microarray production

(Left panel): The ORFeome plasmid collection, which is available in Gateway entry vectors, is transferred to prey destination vectors through recombinatorial cloning. The resulting prey plasmids are individually mixed with a fluorescence reporter plasmid, transfection reagent and a number of additional agents. These reverse transfection mixes are then printed on polystyrene plates, dried and stored until needed for screening. Microarray screening (right panel): Cells are transfected in bulk with the desired bait fusion plasmid (MAPPIT) or the DHFR anchor plasmid (MASPIT) and added to the microarray plates, and the cognate cytokine is added to the appropriate wells, depending on the stimulation schedule. One day later, cells are treated with the appropriate stimuli: depending on the assay (MAPPIT or MASPIT) and the experimental setup (screen in standard conditions or comparing two physiological states), wells are either left untreated or treated with cytokine, fusion compound or additional stimuli. Next, cells are grown for another 2 days until the plates are scanned using a fluorescence imager. (Figure and description from [185])

II.3.5.4 Interactomics and MAPPIT

The MAPPIT technology has been applied in a large-scale interactome mapping program led by the Centre for Cancer Systems Biology (CCSB) located at the Dana Farber Cancer Institute in Boston (US). This "Human Interactome Project" has as main goal the generation of the first reference map of the human interactome network. Their strategy involves mapping PPIs via Y2H assays and confirming the interactions found using alternative binary assays like MAPPIT.

Two proteome-scale studies have been reported on so far. The first one (HI-I-05) was performed by Rual *et al.* in 2004[158]: this study covering ~7.000 genes lead to the identification of 2.700 binary interactions, using a stringent, high-throughput yeast two-hybrid system. The second study (HI-II-14)[186] identified ~14.000 binary interactions starting from combinations of proteins encoded by 13.000 genes.

The MAPPIT system was applied as a validation assay to confirm the interactions of the HI-II-14 study[186]. In the same study two other methods have been used to confirm the results: the in vitro nucleic acid programmable protein array (wNAPPA)[177] assay, and a protein-fragment complementation assay (PCA)[187]. In all cases, the amount of interactions detected is dependent on the chosen threshold as shown in **Figure 41**. The percentage of interactions detected for each different threshold is similar to the one obtained from a tested positive reference set formed by known interactions, described in literature and databases, and clearly above a random reference set (randomly chosen protein pairs) that were tested. Overall, MAPPIT performance was found to be superior.

Similarly, the MAPPIT technology was also used to validate results for smaller systematic interactome mapping experiments: in 2008 for yeast[156], and in 2009 for C. elegans[188].

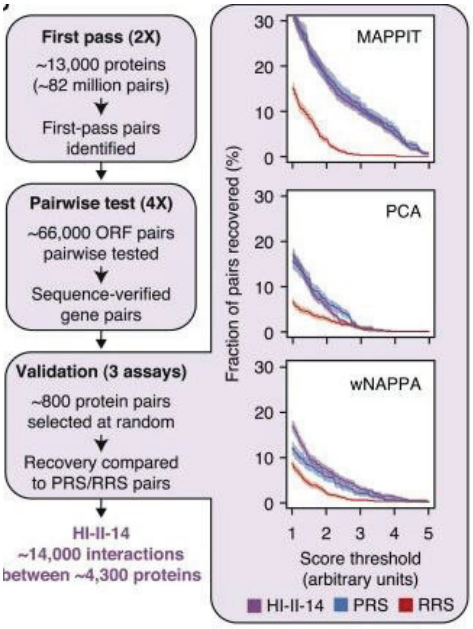


Figure 41 MAPPIT, PCA and wNAPPA to detect PPIs.
The graph on the left shows fraction of HI-II-14, PRS (Positive Reference Set), and RRS (Random reference Set) pairs (right) recovered by MAPPIT, PCA, and wNAPPA at increasing assay stringency. Figure from[186]

Yu and colleagues[156] described a high quality binary protein interaction map of the yeast interactome network: their final dataset ("CCSB-YI1") included 1809 interactions among 1278 proteins.PCA and MAPPIT technology were successfully used to analyze the quality level of the dataset: 94 randomly chosen interaction were retested for confirmation. The following graph shows the results in percentage of positive interactions detected:

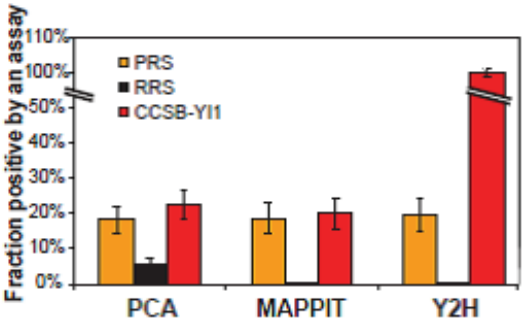
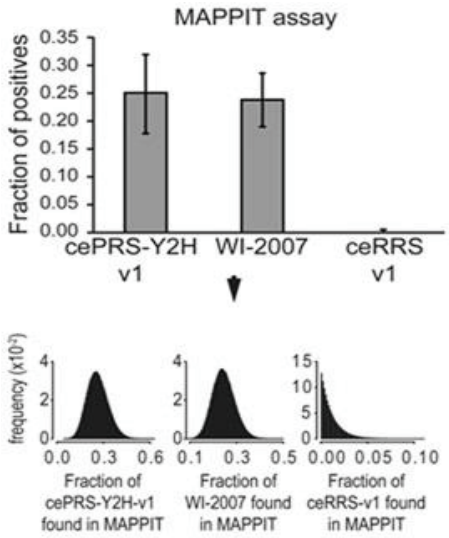


Figure 42 Fraction of protein pairs in PRS, RRS and in the dataset CCSB-YI1 that are positive by PCA, MAPPIT and Y2H.
Picture adapted from [156].



The same type of study has was carried out by Simonis and colleague in 2009[188] to investigate the *C. elegans* interaction network. The study allowed to test a matrix of ~10.000 * ~10.000 proteins and produced a new dataset called WI2007, which includes 1816 PPIs between 1496 proteins. To characterize the new dataset, a novel framework developed by Venkatesan *et al*. was applied[189]. Samples from the dataset (WI-2007) have been used and compared with a PRS (cePRS-v1) and a RRS (ceRRS-v1) dataset in MAPPIT experiment. Results are shown in **Figure 43.**

.

Figure 43 Characterization of WI-2007 database
The proportions of cePRS-Y2H-v1, a random sample of WI-2007 and ceRRS-v1 observed in MAPPIT; the fraction of positive found for WI-2007 is comparable with the positive dataset and significantly higher than the random dataset fraction. Picture adapted from [188].

II.3.6 Quality of the data

The methods explained above obviously produce large amounts of data. Different techniques offer different levels of false positives and false negatives and a specific coverage of the proteome. That is why defining the quality of the data becomes really important. As mentioned earlier, in 2009, Venkatesan and colleagues proposed a method to systematically evaluate the quality of individual binary PPIs reported in interaction mapping experiments[189]. The interaction mapping framework developed by these researchers tries to consider every source of false negatives and false positives and includes four parameters to define the quality: screening completeness, assay sensitivity, sampling sensitivity and precision (**Figure 44**). A subset of data is retested in orthogonal assays considering the rate of false positives and false negatives of each method. For this aim two subsets are used: a positive one (PRS) and a random reference set (RRS) consisting of well documented pairs of interacting human proteins and randomly chosen protein pairs, respectively. As already mentioned, an interaction tool kit was introduced by Braun and colleagues[177] consisting of four complementary high-throughput protein interaction assays: PCA, LUMIER, NAPPA and MAPPIT. After a screening, all the positive interactions are retested in the four assays. In this way data resulting from an Y2H based proteome-wide screen are retested extensively.

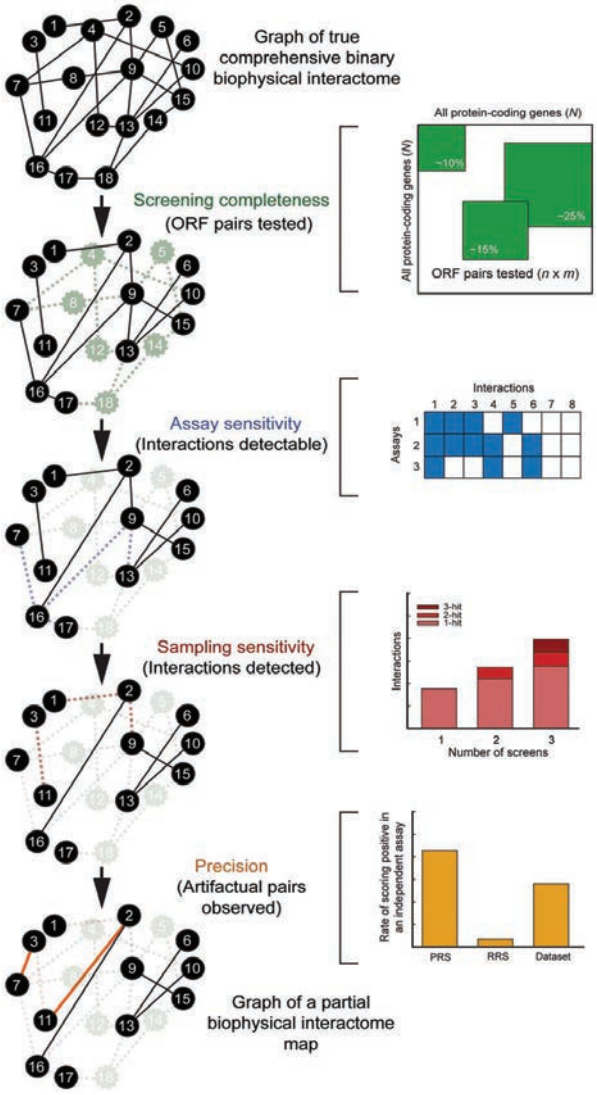


Figure 44 Conceptual framework for interactome mapping.
Screening completeness (fraction of all pairwise protein combinations tested), assay sensitivity (fraction of all biophysical interactions identifiable by a given assay), sampling sensitivity (fraction of all identifiable interactions that are detected in a single trial) and precision (fraction of pairs reported by a given assay that are true positives) are first independently studied and then combined. PRS: positive reference set; RRS: random reference set. Solid black lines are true biophysical interactions, dashed lines are true biophysical interactions missing in that network, solid colored lines are biophysical artefactual pairs .Figure from [189].

II.3.7 PPI databases

Many databases for PPIs have been recently created to be able to merge and unify the enormous amount of data and information coming from high throughput experiments and computational analysis. The databases can be directed towards specific subsets of interactions e.g., kinase–substrate interactions, or based on a specific organism or disease.

Szklarczyk and Jensen classified the existing databases in 3 categories [190]. The first is for pathway databases where the consensus knowledges are collected by experts, like Reactome[191] and KEGG[192]. Reactome[193] is a free source database, curated by PhDs, that assures a high specificity: information is detailed and precise for each interaction, including stoichiometry of the reaction, type of interaction, substrates and localization, and roles played in different diseases. Recently, collaboration with WikiPathways (a community-based platform) started: data of Reactome pathways are converted to the GPML format used in WikiPathways enlarging the coverage of the database showing the importance of the collaboration between the different resources[194]. Kyoto Encyclopedia of Genes and Genomes or KEGG[195], is one of the most widely used pathway databases because it includes the largest number of different species (about 1.500 genomes). It allows linkage with diseases and drugs where a disease is the perturbed status of the molecular system and drugs are perturbants of the molecular system. All the information is offered in pathway maps or in molecule lists [192]. Also the Gene Ontology (GO) Consortium operated towards the same aim of developing tools to unify the different information. They created vocabularies (ontologies) which can be applied to all eukaryotes: biological process, molecular function and cellular component[184]. Also the Sequence Ontology was included to permit the classification and standard representation of sequence features[196]. The second type of databases includes those collecting primary experimental data, for example BioGRID. BioGRID (The Biological General Repository for Interaction Datasets) is one of the largest databases experimentally verified containing 749.912 interactions as drawn from 43.149 publications that represent 30 model organisms[197]. Beside PPIs, chemical and genetic ones, the tool includes data on the influence of protein post-translational modifications, such as phosphorylation and ubiquitination. MINT (The Molecular INTeraction database) is another excellent example of this kind of databases: only physical interactions are included while genetic or computationally identified interactions are not[198,199]. Finally, the last type of databases includes computationally predicted interactors, but performs no manual curation. STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) for example is a free source which combines experimental data, computational

prediction methods and public text collections. The new version 10.0 of STRING includes more than 2000 organisms. Physical and functional interactions on a global scale are included in the collection.

A powerful feature of PPI databases is the graphic visualization of the interactions and the pathways instead of offering a list of partners. There are different visualization methods: an example for the JAK1 interaction is shown in **Figure 45**.

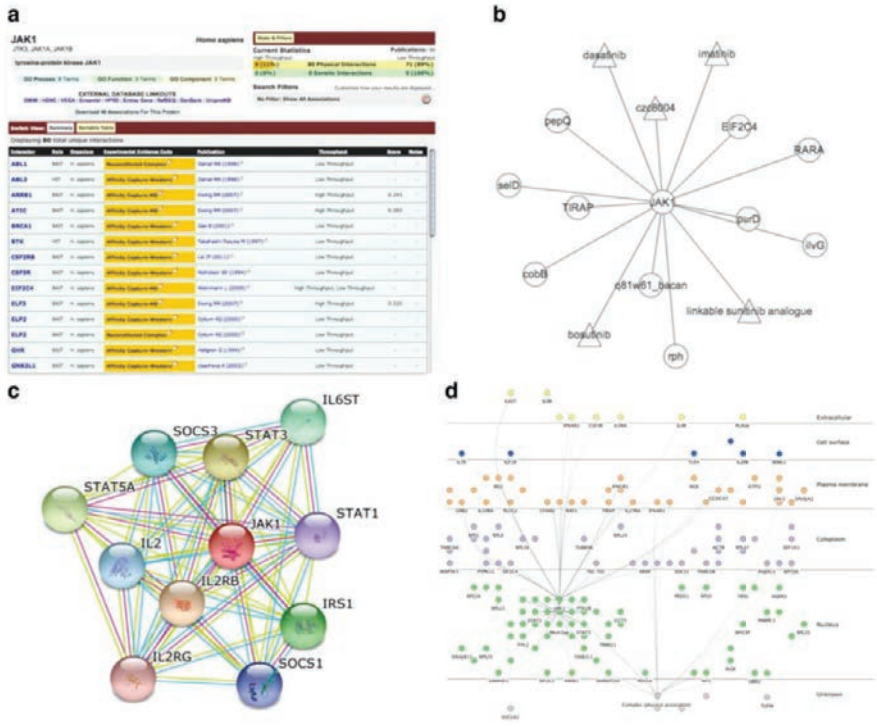


Figure 45 Different databases offer different visualization.
(a) BioGRID HTML table. Physical interactions are shown in yellow as "experimental evidence". (b) JAK1 interactions shown in the IntAct "graph viewer" with small molecules being depicted as triangles. The image shows only the interaction with JAK1and not between other molecules (nodes). (c) JAK1 STRING network. Each different colored line that connects proteins indicates a separate evidence channel for the particular interaction, such as text mining (green), experiments (magenta), and databases (blue). (d) InnateDB's Cytoscape with Cerebral plug-in view uses "cellular component" GO annotation to visualize protein considering their localization. The horizontal lines indicate different cellular compartments. (Figure from [190]*).*

II.4 References

1. Nurse, P. The great ideas of biology. *Clin. Med.* **3,** 560–8 (2003).

2. Vidal, M. A unifying view of 21st century systems biology. *FEBS Lett.* **583,** 3891–4 (2009).

3. Shanahan, T. Francisco J. Ayala and Robert Arp (eds.): Contemporary Debates in Philosophy of Biology. *Sci. Educ.* **19,** 1029–1034 (2010).

4. Nurse, P. & Hayles, J. The cell in an era of systems biology. *Cell* **144,** 850–4 (2011).

5. Robinson, C. V, Sali, A. & Baumeister, W. The molecular sociology of the cell. *Nature* **450,** 973–82 (2007).

6. Vidal, M., Cusick, M. E. & Barabási, A.-L. Interactome networks and human disease. *Cell* **144,** 986–98 (2011).

7. Eisenberg, D., Marcotte, E. M., Xenarios, I. & Yeates, T. O. Protein function in the post-genomic era. *Nature* **405,** 823–6 (2000).

8. Oliver, S. Guilt-by-association goes global. *Nature* **403,** 601–3 (2000).

9. Bork, P. *et al.* Protein interaction networks from yeast to human. *Curr. Opin. Struct. Biol.* **14,** 292–9 (2004).

10. Bowers, P. M. *et al.* Prolinks: a database of protein functional linkages derived from coevolution. *Genome Biol.* **5,** R35 (2004).

11. Nooren, I. M. A. & Thornton, J. M. Diversity of protein-protein interactions. *EMBO J.* **22,** 3486–92 (2003).

12. Rao, V. S., Srinivas, K., Sujini, G. N. & Kumar, G. N. S. Protein-protein interaction detection: methods and analysis. *Int. J. Proteomics* **2014,** 147648 (2014).

13. Alexander Golks, D. G. The O-linked N-acetylglucosamine modification in cellular signalling and the immune system. 'Protein Modifications: Beyond the Usual Suspects' Review Series. *EMBO Rep.* **9,** 748 (2008).

14. Hoffman, M. D., Sniatynski, M. J. & Kast, J. Current approaches for global post-translational modification discovery and mass spectrometric analysis. *Anal. Chim. Acta* **627,** 50–61 (2008).

15. Janke, C., Rogowski, K. & van Dijk, J. Polyglutamylation: a fine-regulator of protein function? 'Protein Modifications: beyond the usual suspects' review series. *EMBO Rep.* **9,** 636–41 (2008).

16. Krueger, K. E. & Srivastava, S. Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics. *Mol. Cell. Proteomics* **5,** 1799–810 (2006).

17. Rogers, L. D. & Overall, C. M. Proteolytic post-translational modification of proteins: proteomic tools and methodology. *Mol. Cell. Proteomics* **12,** 3532–42 (2013).

18. Giglione, C., Boularot, A. & Meinnel, T. Protein N-terminal methionine excision. *Cell. Mol. Life*

Sci. **61,** 1455–1474 (2004).

19. Trewavas, A. Post-Translational Modification of Proteins by Phosphorylation. *Annu. Rev. Plant Physiol.* **27,** 349–374 (1976).

20. Nardozzi, J. D. *et al.* Phosphorylation meets nuclear import: a review. *Cell Commun. Signal.* **8,** 32 (2010).

21. Humphrey, S. J. *et al.* Protein Phosphorylation: A Major Switch Mechanism for Metabolic Regulation. *Trends Endocrinol. Metab.* **26,** 676–687 (2015).

22. Khoury, G. A. *et al.* Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1,** 4–8 (2011).

23. Miranda, M. & Sorkin, A. Regulation of receptors and transporters by ubiquitination: new insights into surprisingly similar mechanisms. *Mol. Interv.* **7,** 157–67 (2007).

24. Komander, D. The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* **37,** 937–53 (2009).

25. Pickart, C. M. Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.* **70,** 503–33 (2001).

26. Amerik, A. Y. & Hochstrasser, M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta* **1695,** 189–207 (2004).

27. Biggar, K. K. & Li, S. S.-C. Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell Biol.* **16,** 5–17 (2014).

28. Bedford, M. T. in *Encycl. Ref. Genomics Proteomics Mol. Med.* 1070–1075 (Springer Berlin Heidelberg, 2006). doi:10.1007/3-540-29623-9_2780

29. Glozak, M. A., Sengupta, N., Zhang, X. & Seto, E. Acetylation and deacetylation of non-histone proteins. *Gene* **363,** 15–23 (2005).

30. Yang, X.-J. & Seto, E. Lysine acetylation: codified crosstalk with other posttranslational modifications. *Mol. Cell* **31,** 449–61 (2008).

31. Verdin, E. & Ott, M. 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat. Rev. Mol. Cell Biol.* **16,** 258–264 (2014).

32. Grangeasse, C., StÃ¼lke, Jã¶. & Mijakovic, I. Regulatory potential of post-translational modifications in bacteria. *Front. Microbiol.* **6,** 500 (2015).

33. Stynen, B., Tournu, H., Tavernier, J. & Van Dijck, P. Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. *Microbiol. Mol. Biol. Rev.* **76,** 331–82 (2012).

34. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340,** 245–6 (1989).

35. Keegan, L., Gill, G. & Ptashne, M. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science (80-.).* **231,** 699–704 (1986).

36. Ma, J. & Ptashne, M. Converting a eukaryotic transcriptional inhibitor into an activator. *Cell*

55, 443–446 (1988).

37. Fields, S. & Sternglanz, R. The two-hybrid system: an assay for protein-protein interactions. *Trends Genet.* **10,** 286–92 (1994).

38. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403,** 623–7 (2000).

39. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 4569–74 (2001).

40. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **10,** 2763–88 (2009).

41. Hirst, M. *et al.* A two-hybrid system for transactivator bait proteins. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 8726–31 (2001).

42. Huang, A. *et al.* Identification of a novel c-Myc protein interactor, JPO2, with transforming activity in medulloblastoma cells. *Cancer Res.* **65,** 5607–19 (2005).

43. Wafa, L. A. *et al.* Isolation and identification of L-dopa decarboxylase as a protein that binds to and enhances transcriptional activity of the androgen receptor using the repressed transactivator yeast two-hybrid system. *Biochem. J.* **375,** 373–83 (2003).

44. Johnsson, N. & Varshavsky, A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 10340–4 (1994).

45. Aronheim, A. Improved efficiency sos recruitment system: expression of the mammalian GAP reduces isolation of Ras GTPase false positives. *Nucleic Acids Res.* **25,** 3373–4 (1997).

46. Medici, R., Bianchi, E., Di Segni, G. & Tocchini-Valentini, G. P. Efficient signal transduction by a chimeric yeast-mammalian G protein alpha subunit Gpa1-Gsalpha covalently fused to the yeast receptor Ste2. *EMBO J.* **16,** 7241–9 (1997).

47. Suter, B., Kittanakom, S. & Stagljar, I. Two-hybrid technologies in proteomics research. *Curr. Opin. Biotechnol.* **19,** 316–23 (2008).

48. Vidal, M., Brachmann, R. K., Fattaey, A., Harlow, E. & Boeke, J. D. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci.* **93,** 10315–10320 (1996).

49. Licitra, E. J. & Liu, J. O. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc. Natl. Acad. Sci.* **93,** 12817–12821 (1996).

50. Broder, Y. C., Katz, S. & Aronheim, A. The ras recruitment system, a novel approach to the study of protein-protein interactions. *Curr. Biol.* **8,** 1121–4 (1998).

51. Gunde, T. & Barberis, A. Yeast growth selection system for detecting activity and inhibition of dimerization-dependent receptor tyrosine kinase. *Biotechniques* **39,** 541–9 (2005).

52. Aronheim, A., Zandi, E., Hennemann, H., Elledge, S. J. & Karin, M. Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Mol. Cell. Biol.* **17,** 3094–102 (1997).

53. Stagljar, I., Korostensky, C., Johnsson, N. & te Heesen, S. A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5187–92 (1998).

54. Thaminy, S., Auerbach, D., Arnoldo, A. & Stagljar, I. Identification of novel ErbB3-interacting factors using the split-ubiquitin membrane yeast two-hybrid system. *Genome Res.* **13,** 1744–53 (2003).

55. Serebriiskii, I. G. *et al.* Detection of peptides, proteins, and drugs that selectively interact with protein targets. *Genome Res.* **12,** 1785–91 (2002).

56. Ehrhard, K. N., Jacoby, J. J., Fu, X. Y., Jahn, R. & Dohlman, H. G. Use of G-protein fusions to monitor integral membrane protein-protein interactions in yeast. *Nat. Biotechnol.* **18,** 1075–9 (2000).

57. Petrascheck, M., Castagna, F. & Barberis, A. Two-hybrid selection assay to identify proteins interacting with polymerase II transcription factors and regulators. *Biotechniques* **30,** 296–8, 300, 302 (2001).

58. Hubsman, M., Yudkovsky, G. & Aronheim, A. A novel approach for the identification of protein-protein interaction with integral membrane proteins. *Nucleic Acids Res.* **29,** E18 (2001).

59. Urech, D. M., Lichtlen, P. & Barberis, A. Cell growth selection system to detect extracellular and transmembrane protein interactions. *Biochim. Biophys. Acta* **1622,** 117–27 (2003).

60. Tafelmeyer, P., Johnsson, N. & Johnsson, K. Transforming a (beta/alpha)8--barrel enzyme into a split-protein sensor through directed evolution. *Chem. Biol.* **11,** 681–9 (2004).

61. Möckli, N. *et al.* Yeast split-ubiquitin-based cytosolic screening system to detect interactions between transcriptionally active proteins. *Biotechniques* **42,** 725–30 (2007).

62. Dmitrova, M. *et al.* A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in Escherichia coli. *Mol. Gen. Genet.* **257,** 205–12 (1998).

63. Dove, S. L., Joung, J. K. & Hochschild, A. Activation of prokaryotic transcription through arbitrary protein-protein contacts. *Nature* **386,** 627–30 (1997).

64. Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7382–7 (2000).

65. Wang, Z.-H. *et al.* A transferable heterogeneous two-hybrid system in Escherichia coli based on polyhydroxyalkanoates synthesis regulatory protein PhaR. *Microb. Cell Fact.* **10,** 21 (2011).

66. Ding, Z. *et al.* A novel cytology-based, two-hybrid screen for bacteria applied to protein-protein interaction studies of a type IV secretion system. *J. Bacteriol.* **184,** 5572–82 (2002).

67. Kolmar, H. *et al.* Dimerization of Bence Jones proteins: linking the rate of transcription from an Escherichia coli promoter to the association constant of REIV. *Biol. Chem. Hoppe. Seyler.* **375,** 61–70 (1994).

68. Strauch, E.-M. & Georgiou, G. A bacterial two-hybrid system based on the twin-arginine

transporter pathway of E. coli. *Protein Sci.* **16,** 1001–8 (2007).

69. Hao, N., Whitelaw, M. L., Shearwin, K. E., Dodd, I. B. & Chapman-Smith, A. Identification of residues in the N-terminal PAS domains important for dimerization of Arnt and AhR. *Nucleic Acids Res.* **39,** 3695–709 (2011).

70. Ozawa, T., Takeuchi, M., Kaihara, A., Sato, M. & Umezawa, Y. Protein Splicing-Based Reconstitution of Split Green Fluorescent Protein for Monitoring Protein−Protein Interactions in Bacteria: Improved Sensitivity and Reduced Screening Time. *Anal. Chem.* **73,** 5866–5874 (2001).

71. Feng, X. H. & Derynck, R. Mammalian two-hybrid assays. Analyzing protein-protein interactions in transforming growth factor-beta signaling pathway. *Methods Mol. Biol.* **177,** 221–39 (2001).

72. Dang, C. V *et al.* Intracellular leucine zipper interactions suggest c-Myc hetero-oligomerization. *Mol. Cell. Biol.* **11,** 954–62 (1991).

73. Maroun, M. & Aronheim, A. A novel in vivo assay for the analysis of protein-protein interaction. *Nucleic Acids Res.* **27,** i–v (1999).

74. Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I. & Stefan, E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat. Rev. Drug Discov.* **6,** 569–82 (2007).

75. Singh, K., Kang, P. J. & Park, H.-O. The Rho5 GTPase is necessary for oxidant-induced cell death in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 1522–7 (2008).

76. Miller, K. E., Kim, Y., Huh, W.-K. & Park, H.-O. Bimolecular Fluorescence Complementation (BiFC) Analysis: Advances and Recent Applications for Genome-Wide Interaction Studies. *J. Mol. Biol.* **427,** 2039–55 (2015).

77. Morell, M., Ventura, S. & Avilés, F. X. Protein complementation assays: approaches for the in vivo analysis of protein interactions. *FEBS Lett.* **583,** 1684–91 (2009).

78. Villalobos, V., Naik, S. & Piwnica-Worms, D. Current state of imaging protein-protein interactions in vivo with genetically encoded reporters. *Annu. Rev. Biomed. Eng.* **9,** 321–49 (2007).

79. Ghosh, I., Hamilton, A. D. & Regan, L. Antiparallel Leucine Zipper-Directed Protein Reassembly: Application to the Green Fluorescent Protein. *J. Am. Chem. Soc.* **122,** 5658–5659 (2000).

80. Kerppola, T. K. Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nat. Protoc.* **1,** 1278–1286 (2006).

81. Kerppola, T. K. Bimolecular fluorescence complementation (BiFC) analysis of protein interactions in live cells. *Cold Spring Harb. Protoc.* **2013,** 727–31 (2013).

82. Uyttendaele, I. MAPPIT analysis of medically relevant protein protein interactions, PhD thesis. (2012).

83. Kelly, K. F. & Daniel, J. M. POZ for effect--POZ-ZF transcription factors in cancer and development. *Trends Cell Biol.* **16,** 578–87 (2006).

84. Melnick, A. *et al.* In-depth mutational analysis of the promyelocytic leukemia zinc finger BTB/POZ domain reveals motifs and residues required for biological and transcriptional functions. *Mol. Cell. Biol.* **20,** 6550–67 (2000).

85. Ahmad, K. F. *et al.* Mechanism of SMRT corepressor recruitment by the BCL6 BTB domain. *Mol. Cell* **12,** 1551–64 (2003).

86. Melnick, A. *et al.* Critical residues within the BTB domain of PLZF and Bcl-6 modulate interaction with corepressors. *Mol. Cell. Biol.* **22,** 1804–18 (2002).

87. Huynh, K. D. & Bardwell, V. J. The BCL-6 POZ domain and other POZ domains interact with the co-repressors N-CoR and SMRT. *Oncogene* **17,** 2473–84 (1998).

88. Deltour, S., Guerardel, C. & Leprince, D. Recruitment of SMRT/N-CoR-mSin3A-HDAC-repressing complexes is not a general mechanism for BTB/POZ transcriptional repressors: the case of HIC-1 and gammaFBP-B. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14831–6 (1999).

89. Peukert, K. *et al.* An alternative pathway for gene regulation by Myc. *EMBO J.* **16,** 5672–86 (1997).

90. Petschnigg, J. *et al.* The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. *Nat. Methods* **11,** 585–92 (2014).

91. Förster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.* **437,** 55–75 (1948).

92. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 719–26 (1967).

93. Broussard, J. A., Rappaz, B., Webb, D. J. & Brown, C. M. Fluorescence resonance energy transfer microscopy as demonstrated by measuring the activation of the serine/threonine kinase Akt. *Nat. Protoc.* **8,** 265–81 (2013).

94. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*. (1983). at <http://books.google.be/books/about/Principles_of_Fluorescence_Spectroscopy.html?id=moB8QgAACAAJ&pgis=1>

95. Piston, D. W. & Kremers, G.-J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends Biochem. Sci.* **32,** 407–14 (2007).

96. George Abraham, B. *et al.* Fluorescent Protein Based FRET Pairs with Improved Dynamic Range for Fluorescence Lifetime Measurements. *PLoS One* **10,** e0134436 (2015).

97. Xu, Y., Piston, D. W. & Johnson, C. H. A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 151–6 (1999).

98. Eyckerman, S. *et al.* Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell Biol.* **3,** 1114–9 (2001).

99. Tavernier, J. *et al.* MAPPIT: a cytokine receptor-based two-hybrid method in mammalian

cells. *Clin Exp Allergy* **32,** 1397–1404 (2002).

100. Lavens, D. *et al.* A complex interaction pattern of CIS and SOCS2 with the leptin receptor. *J. Cell Sci.* **119,** 2214–24 (2006).

101. Montoye, T. *et al.* Analysis of leptin signalling in hematopoietic cells using an adapted MAPPIT strategy. *FEBS Lett.* **580,** 3301–7 (2006).

102. Lemmens, I. *et al.* Heteromeric MAPPIT: a novel strategy to study modification-dependent protein-protein interactions in mammalian cells. *Nucleic Acids Res.* **31,** e75 (2003).

103. Lievens, S., Peelman, F., De Bosscher, K., Lemmens, I. & Tavernier, J. MAPPIT: a protein interaction toolbox built on insights in cytokine receptor signaling. *Cytokine Growth Factor Rev.* **22,** 321–9 (2011).

104. Caligiuri, M. *et al.* MASPIT: three-hybrid trap for quantitative proteome fingerprinting of small molecule-protein interactions in mammalian cells. *Chem. Biol.* **13,** 711–22 (2006).

105. Ulrichts, P., Peelman, F., Beyaert, R. & Tavernier, J. MAPPIT analysis of TLR adaptor complexes. *FEBS Lett* **581,** 629–636 (2007).

106. Ulrichts, P. & Tavernier, J. MAPPIT analysis of early Toll-like receptor signalling events. *Immunol. Lett.* **116,** 141–8 (2008).

107. Eyckerman, S. *et al.* Reverse MAPPIT: screening for protein-protein interaction modifiers in mammalian cells. *Nat. Methods* **2,** 427–33 (2005).

108. Sasaki, M. *et al.* Regulation of the MDM2-P53 pathway and tumor growth by PICT1 via nucleolar RPL11. *Nat. Med.* **17,** 944–51 (2011).

109. Vassilev, L. T. *et al.* In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303,** 844–8 (2004).

110. Wisniewski, D. *et al.* Characterization of potent inhibitors of the Bcr-Abl and the c-kit receptor tyrosine kinases. *Cancer Res.* **62,** 4244–55 (2002).

111. Lievens, S. *et al.* KISS, a mammalian in situ protein interaction sensor. *Mol. Cell. Proteomics* **13,** 3332–42 (2014).

112. de Souza, N. Biochemistry: Protein interactions in situ. *Nat. Methods* **11,** 1093–1093 (2014).

113. Kaboord, B. & Perr, M. Isolation of proteins and protein complexes by immunoprecipitation. *Methods Mol. Biol.* **424,** 349–64 (2008).

114. Golemis, E. *Protein-protein interactions : a molecular cloning manual*. (Cold Spring Harbor Laboratory Press, 2002).

115. Phizicky, E. M. & Fields, S. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59,** 94–123 (1995).

116. Barrios-Rodiles, M. *et al.* High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* **307,** 1621–5 (2005).

117. Taipale, M. *et al.* A quantitative chaperone interaction network reveals the architecture of

cellular protein homeostasis pathways. *Cell* **158,** 434–48 (2014).

118. Taipale, M. *et al.* Quantitative analysis of HSP90-client interactions reveals principles of substrate recognition. *Cell* **150,** 987–1001 (2012).

119. Trepte, P. *et al.* DULIP: A Dual Luminescence-Based Co-Immunoprecipitation Assay for Interactome Mapping in Mammalian Cells. *J. Mol. Biol.* **427,** 3375–88 (2015).

120. Buntru, A., Trepte, P., Klockmeier, K., Schnoegl, S. & Wanker, E. E. Current Approaches Toward Quantitative Mapping of the Interactome. *Front. Genet.* **7,** 74 (2016).

121. Terpe, K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60,** 523–33 (2003).

122. Ratner, D. The interaction of bacterial and phage proteins with immobilized Escherichia coli RNA polymerase. *J. Mol. Biol.* **88,** 373–383 (1974).

123. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nat. Biotechnol.* **17,** 1030–2 (1999).

124. Günzl, A. & Schimanski, B. Tandem affinity purification of proteins. *Curr. Protoc. Protein Sci.* **Chapter 19,** Unit 19.19 (2009).

125. Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24,** 218–29 (2001).

126. Völkel, P., Le Faou, P. & Angrand, P.-O. Interaction proteomics: characterization of protein complexes using tandem affinity purification-mass spectrometry. *Biochem. Soc. Trans.* **38,** 883–7 (2010).

127. Kaake, R. M., Wang, X. & Huang, L. Profiling of protein interaction networks of protein complexes using affinity purification and quantitative mass spectrometry. *Mol. Cell. Proteomics* **9,** 1650–65 (2010).

128. Young, C. L., Britton, Z. T. & Robinson, A. S. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol. J.* **7,** 620–634 (2012).

129. Wang, X., Campoli, M., Ko, E., Luo, W. & Ferrone, S. Enhancement of scFv fragment reactivity with target antigens in binding assays following mixing with anti-tag monoclonal antibodies. *J. Immunol. Methods* **294,** 23–35 (2004).

130. Evan, G. I., Lewis, G. K., Ramsay, G. & Bishop, J. M. Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol. Cell. Biol.* **5,** 3610–6 (1985).

131. McKern, N. M. *et al.* Crystallization of the first three domains of the human insulin-like growth factor-1 receptor. *Protein Sci.* **6,** 2663–6 (1997).

132. Hopp, T. P. *et al.* A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology* **6,** 1204–1210 (1988).

133. The Human Protein Atlas. at <http://www.proteinatlas.org>

134. Söderberg, O. *et al.* Direct observation of individual endogenous protein complexes in situ by

proximity ligation. *Nat. Methods* **3,** 995–1000 (2006).

135. Ali, M. M. *et al.* Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chem. Soc. Rev.* **43,** 3324–41 (2014).

136. Clausson, C.-M. *et al.* Compaction of rolling circle amplification products increases signal integrity and signal-to-noise ratio. *Sci. Rep.* **5,** 12317 (2015).

137. Lundberg, M. *et al.* Multiplexed homogeneous proximity ligation assays for high-throughput protein biomarker research in serological material. *Mol. Cell. Proteomics* **10,** M110.004978 (2011).

138. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293,** 2101–5 (2001).

139. Hall, D. A., Ptacek, J. & Snyder, M. Protein microarray technology. *Mech. Ageing Dev.* **128,** 161–7 (2007).

140. Bertone, P. & Snyder, M. Advances in functional protein microarray technology. *FEBS J.* **272,** 5400–11 (2005).

141. Speer, R., Wulfkuhle, J. D., Liotta, L. A. & Petricoin, E. F. Reverse-phase protein microarrays for tissue-based analysis. *Curr. Opin. Mol. Ther.* **7,** 240–5 (2005).

142. Orlando, V. Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **25,** 99–104 (2000).

143. Kluger, R. & Alagic, A. Chemical cross-linking and protein-protein interactions-a review with illustrative protocols. *Bioorg. Chem.* **32,** 451–72 (2004).

144. Vasilescu, J., Guo, X. & Kast, J. Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* **4,** 3845–54 (2004).

145. Suchanek, M., Radzikowska, A. & Thiele, C. Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat. Methods* **2,** 261–7 (2005).

146. Rees, J. S., Li, X.-W., Perrett, S., Lilley, K. S. & Jackson, A. P. Protein Neighbors and Proximity Proteomics. *Mol. Cell. Proteomics* **14,** 2848–2856 (2015).

147. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196,** 801–10 (2012).

148. Roux, K. J., Kim, D. I. & Burke, B. BioID: a screen for protein-protein interactions. *Curr. Protoc. Protein Sci.* **74,** Unit 19.23. (2013).

149. Firat-Karalar, E. N. & Stearns, T. Probing mammalian centrosome structure using BioID proximity-dependent biotinylation. *Methods Cell Biol.* **129,** 153–70 (2015).

150. Kim, D. I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* mbc.E15–12–0844– (2016). doi:10.1091/mbc.E15-12-0844

151. Bartel, P. L., Roecklein, J. A., SenGupta, D. & Fields, S. A protein linkage map of Escherichia coli bacteriophage T7. *Nat. Genet.* **12,** 72–7 (1996).

152. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* **16,** 277–82 (1997).

153. Walhout, A. J. Protein Interaction Mapping in C. elegans Using Proteins Involved in Vulval Development. *Science (80-.).* **287,** 116–122 (2000).

154. Reboul, J. *et al.* C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat. Genet.* **34,** 35–41 (2003).

155. Giot, L. *et al.* A protein interaction map of Drosophila melanogaster. *Science* **302,** 1727–36 (2003).

156. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322,** 104–10 (2008).

157. Walhout, A. J. M. *et al. Applications of Chimeric Genes and Hybrid Proteins - Part C: Protein-Protein Interactions and Genomics*. *Methods Enzymol.* **328,** (Elsevier, 2000).

158. Rual, J.-F. *et al.* Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14,** 2128–35 (2004).

159. Strausberg, R. L. *et al.* Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 16899–903 (2002).

160. Kachroo, A. H. *et al.* Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* **348,** 921–5 (2015).

161. Laurent, J. M., Young, J. H., Kachroo, A. H. & Marcotte, E. M. Efforts to make and apply humanized yeast. *Brief. Funct. Genomics* **15,** 155–63 (2016).

162. Kataoka, T. *et al.* Functional homology of mammalian and yeast RAS genes. *Cell* **40,** 19–26 (1985).

163. Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–7 (2002).

164. Ho, Y. *et al.* Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* **415,** 180–3 (2002).

165. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440,** 637–43 (2006).

166. Gavin, A.-C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440,** 631–6 (2006).

167. Ewing, R. M. *et al.* Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3,** 89 (2007).

168. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162,** 425–440 (2015).

169. Yang, X. *et al.* A public genome-scale lentiviral expression library of human ORFs. *Nat. Methods* **8,** 659–61 (2011).

170. Gingras, A.-C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **8,** 645–54 (2007).

171. Mellacheruvu, D. *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat. Methods* **10,** 730–6 (2013).

172. Sinz, A. Investigation of protein-protein interactions in living cells by chemical crosslinking and mass spectrometry. *Anal. Bioanal. Chem.* **397,** 3433–40 (2010).

173. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6,** 83–90 (2009).

174. Eyckerman, S. *et al.* Trapping mammalian protein complexes in viral particles. *Nat. Commun.* **7,** 11416 (2016).

175. Royer, M., Hong, S. S., Gay, B., Cerutti, M. & Boulanger, P. Expression and extracellular release of human immunodeficiency virus type 1 Gag precursors by recombinant baculovirus-infected cells. *J. Virol.* **66,** 3230–5 (1992).

176. Briggs, J. A. G. *et al.* The stoichiometry of Gag protein in HIV-1. *Nat. Struct. Mol. Biol.* **11,** 672–5 (2004).

177. Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein interactions. *Nat. Methods* **6,** 91–7 (2009).

178. Titeca, K. *et al.* SFINX: Straightforward Filtering Index for Affinity Purification-Mass Spectrometry Data Analysis. *J. Proteome Res.* **15,** 332–8 (2016).

179. Schnell, J. R., Dyson, H. J. & Wright, P. E. Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annu. Rev. Biophys. Biomol. Struct.* **33,** 119–40 (2004).

180. Tarassov, K. *et al.* An in vivo map of the yeast protein interactome. *Science* **320,** 1465–70 (2008).

181. Lievens, S. *et al.* Array MAPPIT: high-throughput interactome analysis in mammalian cells. *J. Proteome Res.* **8,** 877–86 (2009).

182. Lievens, S. *et al.* Design of a fluorescence-activated cell sorting-based Mammalian protein-protein interaction trap. *Methods Mol. Biol.* **263,** 293–310 (2004).

183. Lamesch, P. *et al.* hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* **89,** 307–15 (2007).

184. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–9 (2000).

185. Lievens, S., et al. . Proteome-scale binary interactomics in human cells. *Pap. Submitt.* (2016).

186. Rolland, T. *et al.* A Proteome-Scale Map of the Human Interactome Network. *Cell* **159,** 1212–1226 (2014).

187. Nyfeler, B., Michnick, S. W. & Hauri, H.-P. Capturing protein interactions in the secretory pathway of living cells. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 6350–5 (2005).

188. Simonis, N. *et al.* Empirically controlled mapping of the Caenorhabditis elegans protein-protein interactome network. *Nat. Methods* **6,** 47–54 (2009).

189. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nat. Methods* **6,** 83–90 (2009).

190. Szklarczyk, D. & Jensen, L. J. Protein-protein interaction databases. *Methods Mol. Biol.* **1278,** 39–56 (2015).

191. Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* **39,** D691–7 (2011).

192. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38,** D355–60 (2010).

193. Reactome. at <http://www.reactome.org>

194. Bohler, A. *et al.* Reactome from a WikiPathways Perspective. *PLoS Comput. Biol.* **12,** e1004941 (2016).

195. KEGG. at <http://www.genome.jp/kegg/>

196. Harris, M. A. *et al.* The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32,** 258D–261 (2004).

197. Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* **43,** D470–8 (2015).

198. Chatr-aryamontri, A. *et al.* MINT: the Molecular INTeraction database. *Nucleic Acids Res.* **35,** D572–4 (2007).

199. MINT. at <http://mint.bio.uniroma2.it/mint/>

Scientific problems addressed and Aims of the project

The word “Cancer” (gr. καρκίνωμα, from καρκίνος, crab) was proposed by Hippocrates (460-370 BC) to define uncontrolled growth and proliferation events: he introduced the terms *carcinos* and *carcinoma* to describe ulcer/non-ulcer-forming tumors. The evidence of cancer existence in humans and other animals is dated back to the beginning of recorded history, the first record being the Edwin Smith Papyrus, from a 3000 BC Egyptian text book. Remarkably, here the author stated in that “There is no treatment” for this disease. Despite an increased basic understanding of cancer, in 2016, cancer remains in many cases an untreatable disease and a leading cause of death worldwide (8.2 million deaths in 2012)[1]. Lung cancer and prostate cancer are the most common cancer types in men in developed countries, while breast cancer is the most common and deadliest type of cancer for women. An increase of about 70% of cancer occurrence in the next 20 years has been estimated, mostly due to growth and increase of the average life expectancy of the world’s population, particularly in developing countries.

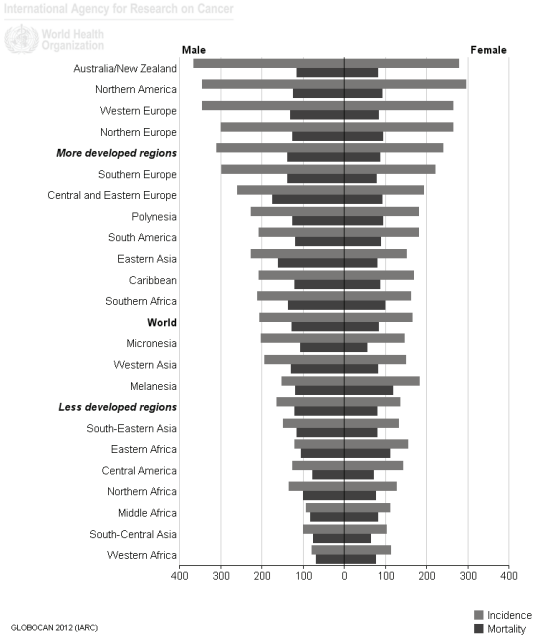


Figure 46 Estimated Incidence, Mortality and Prevalence Worldwide in 2012, All Cancers (excluding non-melanoma skin cancer)

GLOBOCAN is a project which provides information about incidence, mortality and prevalence of different types of cancer, sub-divided in world regions and gender. In 2012, 14.1 million new cancer cases and 8.2 million cancer deaths have been reported. Cancer incidence in the male population is highly variable between different countries while the incidence in the female population is more stable region-wise. Mortality rates are 15% higher in developed world regions for male and 8% for female. (Figure taken from http://www.iarc.fr [2])

Cancer is the results of a combination of genetic and epigenetic alterations and external risk factors. Genetic[3] (mutations or chromosome translocation) and epigenetic[4] (such as histone modifications, and dysregulation of DNA binding proteins) alterations might lead to increased activity of proto-oncogenes and loss of activity of tumor suppressor genes contributing to tumor formation and progression.

Accumulating evidence suggests that non-clustered protocadherins (Pcdhs) can function as tumor suppressors or proto-oncogenes in various non-neuronal human tumours.[5,6]The Pcdh family represents the largest sub-group of the cadherin superfamily [7]. They are mainly but not exclusively expressed in the nervous system. Pcdhs differ in various aspects from classic cadherins and can be subdivided in clustered Pcdhs (C-Pcdhs), and non-clustered Pcdhs (NC-Pcdhs) or δ-protocadherins (δPcdhs)[7,8], based on overall sequence homology, number of EC repeats (seven versus six), and conservation of specific amino acid motifs in cytoplasmic domains. The subfamily of δPcdhs is well conserved among vertebrates and comprises 10 members in humans. δPcdhs can be expressed as short or long isoforms, generated by alternative splicing and differing from each other by the size of their cytoplasmic domain. All the long isoforms show highly conserved motifs (CM1 and CM2 for δ1- and δ2Pcdhs; plus CM3 for δ1Pcdhs only) in their cytoplasmic domains [8].

Evidence from the last few years indicates an important role for δPcdhs in human cancer. For instance, missense mutations were found in PCDH9, -17 and -18 in a large-scale analysis of pancreatic cancer[9]. Pcdh8 is a strong candidate tumor suppressor in breast cancer[10], and also PCDH10 and PCDH20 were proposed as tumor suppressors[11–14]. For PCDH10, literature shows strong evidence for frequent epigenetic inactivation in various human cancers, including nasopharyngeal, gastric and cervical cancer. Pcdhs can also show oncogenic activity, for example PCDH11Y is a candidate proto-oncogene in prostate cancer[15]. More specifically, progression to the hormone-independent status is correlated with expression of a cytoplasmic variant of PCDH11Y, which is the isoform encoded by the male chromosome Y.

Research on the Pcdh family is still in its infancy. In particular, the molecular interaction partners of short or long cytoplasmic domains of δPcdhs are still largely unknown. In general, a preliminary step in understanding protein structure and function is to determine which proteins interact with each other, thereby pointing out relevant biological pathways. The identification of interaction partners might reveal in which signaling pathways Pcdhs are involved and how their functions can be restored or inhibited in function of therapeutic needs. An interesting molecular interaction partner for the

CM3 domain is, for example, protein phosphatase-1-α (PP1α).[8] Specific interaction partners for CM1 or CM2 have not been identified to date.

The aim of the PhD project described in this dissertation has been to investigate the putative tumor suppressor or oncogenic role of δPcdhs by identification and functional analysis of novel intracellular interaction partners and to elucidate their role in different signaling pathways. The focus on the conserved motifs CM1 and CM2 is due to their high evolutionary conservation across species[7], which points at an important role in intracellular signaling. Of course, we cannot exclude *a priori* that this conservation is due to structural needs, i.e. intramolecular interactions rather than intermolecular interactions. Unfortunately, the structure of Pcdh cytoplasmic domains is not yet resolved.

In our studies we focused on PCDH10 as representative of δ2Pcdhs and on PCDH11X as representative for the δ1 subgroup. Other members of the family have been included in various experiments. In the first part of the project we investigated novel molecular interactions of Pcdh cytoplasmic domains using Mammalian Protein-Protein Interaction Trap (MAPPIT) technology. Array MAPPIT was selected as a robust screening assay. Putative interaction partners were confirmed with orthogonal assays and analyzed to obtain a priority list. The aim of the second part of the project was the functional characterization of the molecular interactions discovered, hardly predictable as it depends on the nature of these molecular partners (e.g. novel *versus* well-known proteins, barely studied *versus* well-characterized proteins, structural *versus* regulatory/signaling proteins). The functional implications of these molecular interactions will be studied using contemporary molecular technologies.

References

1. Stewart, B. W. & Wild, C. P. *World cancer report 2014*.

2. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. cancer* **136,** E359–86 (2015).

3. Knudson, A. G. Two genetic hits (more or less) to cancer. *Nat. Rev. Cancer* **1,** 157–62 (2001).

4. Jones, P. A. & Baylin, S. B. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3,** 415–28 (2002).

5. Berx, G. & van Roy, F. Involvement of Members of the Cadherin Superfamily in Cancer. *Cold Spring Harb. Perspect. Biol.* **1,** a003129–a003129 (2009).

6. van Roy, F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nat. Rev. Cancer* **14,** 121–34 (2014).

7. Hulpiau, P. & van Roy, F. Molecular evolution of the cadherin superfamily. *Int. J. Biochem. Cell Biol.* **41,** 349–69 (2009).

8. Vanhalst, K., Kools, P., Staes, K., van Roy, F. & Redies, C. delta-Protocadherins: a gene family expressed differentially in the mouse brain. *Cell Mol Life Sci* **62,** 1247–1259 (2005).

9. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–6 (2008).

10. Yu, J. S. *et al.* PCDH8, the human homolog of PAPC, is a candidate tumor suppressor of breast cancer. *Oncogene* **27,** 4657–65 (2008).

11. Imoto, I. *et al.* Frequent silencing of the candidate tumor suppressor PCDH20 by epigenetic mechanism in non-small-cell lung cancers. *Cancer Res.* **66,** 4617–26 (2006).

12. Narayan, G. *et al.* Protocadherin PCDH10, involved in tumor progression, is a frequent and early target of promoter hypermethylation in cervical cancer. *Genes. Chromosomes Cancer* **48,** 983–92 (2009).

13. Ying, J. *et al.* Functional epigenetics identifies a protocadherin PCDH10 as a candidate tumor suppressor for nasopharyngeal, esophageal and multiple other carcinomas with frequent methylation. *Oncogene* **25,** 1070–80 (2006).

14. Yu, J. *et al.* Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* **136,** 640–51.e1 (2009).

15. Terry, S. *et al.* Protocadherin-PC promotes androgen-independent prostate cancer cell growth. *Prostate* **66,** 1100–13 (2006).

RESULTS AND DISCUSSION

Chapter III. IDENTIFICATION OF NOVEL INTRACELLULAR INTERACTION PARTNERS OF δPCDH FAMILY MEMBERS[c]

c The study was designed by Eleonora Billi, Sam Lievens and Jan Tavernier.
All experiments were done by Eleonora Billi unless stated otherwise.
MAPPIT constructs were generated by Eleonora Billi, Irene Kahr and Uta Brunner.
cDNA library MAPPIT experiments and Y2H tests were designed and carried out by Irene Kahr and Uta Brunner.
The MAPPIT cDNA library screens were performed by Sam Lievens.
The Array MAPPIT experiments were run by Dieter Defever under the supervision of Sam Lievens.

III.1 Introduction

E-cadherin is a transmembrane cell-cell adhesion molecule whose cytoplasmic domain is coupled to the actin cytoskeleton via catenins. We have demonstrated that the E-cadherin/catenin complex plays an instrumental role in tumor suppression[1]. On the other hand, nuclear beta-catenin can act as a potent oncogene. The δ-protocadherin family comprises 10 members, all with an extracellular domain related to E-cadherin but for the rest differing strongly from E-cadherin[2,3]. For instance, their cytoplasmic domains (CDs) occur in at least two forms, which are not interacting with classic catenins. The longer CD type contains two to three conserved motifs (CMs), which most likely serve as interaction domains for signaling proteins. Protocadherins are probably rather communicative than adhesive proteins and functional differences may exist between long and short isoforms. Hence, it is of utmost importance to identify and characterize the still enigmatic cytoplasmic molecular partners of δ-protocadherins (δPcdhs).

δPcdhs are expressed at high levels in the brain. So it is not that surprising that many of them are involved in neurological disorders, such as autism, schizophrenia, epilepsy and Alzheimer's disease. A role in neurological disorders has indeed been reported for PCDH7[4], PCDH10[5], Pcdh8 and Pcdh17[6], PCDH11X and PCDH11Y[7–9] and PCDH18[10], but the most convincing evidence comes from studies on Pcdh19, which is the most relevant gene in epilepsy after SCN1A[11]. PCDHs have also been shown to be involved in other diseases and symptoms, for example, asthma and bronchial hyper responsiveness (BHR)[12]. Recent reports also describe a role for δPcdhs in the immune system[13–16].

Indeed, both direct and indirect lines of evidence indicate that non-clustered (NC) Pcdh genes play important roles during tumorigenesis, either as tumor-suppressor genes or as oncogenes. For instance, missense mutations were found in *Pcdh9, -17 and -18* genes in a large-scale analysis of pancreatic cancer[17]. PCDH8 is a strong candidate tumor suppressor in breast cancer[18] and also PCDH10 and PCDH20 were proposed as tumor suppressors[19–22]. For PCDH10 the literature shows strong evidence for frequent epigenetic inactivation in various human cancers, including nasopharyngeal, gastric and cervical cancer (reviewed by van Roy, 2014, Nat. Rev. Cancer)[23]. On the other hand, there is accumulating evidence for a role of PCDH11 as proto-oncogene in prostate cancer[24]. More specifically, progression to the hormone-independent status of prostate cancer is correlated with expression of a cytoplasmic variant of PCDH11Y, which is an isoform encoded by the male chromosome Y. (More examples and details in Chapter I).

Research on the protocadherin families is still in its infancy. In particular, it is still enigmatic what the molecular interaction partners are of the short or long cytoplasmic domains of δPcdhs. Their identification might reveal in which signaling pathways Pcdhs are involved and how their functions can be restored or inhibited for therapeutic needs. δPcdhs are expressed as long and short isoforms. For long isoforms, the cytoplasmic domains are characterized by the presence of conserved amino acid motifs CM1 to CM3. As these CMs are highly conserved within the δPcdh family, and despite the obvious differences between the members of this family, it is possible and even most likely, that δPcdhs share some core interaction partners.

An interesting molecular interaction partner for CM3 is protein phosphatase-1-α (PP1α)[25]. However, specific interaction partners for CM1 or CM2 have not been reported to date. Nonetheless, several δPcdhs have been shown to interact with intracellular partners and have therefore been implicated in signaling pathways. One of those interactions is the binding of PCDH10 to NAP1 (Nck-associated protein 1), which recruits WAVE1 (SCAR1) to cell-cell contacts, resulting in a non-standard assembly of PCDH10 with F-actin and N-cadherin at this location[26]. In the presence of PCDH10 the complex NAP1-WAVE1 becomes redistributed at cell-cell contacts instead of at the lamellipodia where it is usually localized to participate in the cell migration process. In this way the cell-cell contacts become locally unstable. On the basis of this remarkable interaction, a role for PCDH10 has been proposed in upregulation of the cell motile machine at cell-cell contact sites, by modulating the cadherin-dependent inhibition of cell movement. (For other interaction partners, check Chapter I).

Because δPcdhs are most likely involved in signaling pathways, we were in particular looking for interaction partners that could mediate such signaling. Moreover, the CMs became our major focus, because we speculated that their high evolutionary conservation across species points to an important role in intracellular signaling. Of course, we could not exclude beforehand that this conservation is due to structural needs, i.e. intramolecular interactions rather than intermolecular interactions, since so far structures of Pcdh cytoplasmic domain remain to be resolved.

Different approaches have been followed to identify novel (common or exclusive) interaction partners of δPcdhs and the robust array-MAPPIT technology was chosen to perform the important initial screening experiments.

The Mammalian Protein-Protein Interaction Trap (MAPPIT) is a two hybrid system for the detection of PPIs in intact mammalian cells. The method is described in detail in Eyckerman *et al.*[27] and in Chapter II of this dissertation. Briefly, the technique is based on a modified JAK-STAT signaling pathway. The first hybrid is a modified cytokine receptor fused with the protein of interest and missing STAT3 recruitment sites (the bait). MAPPIT prey constructs are composed of a candidate interaction partner or of an ORF C-terminally fused to a part of the gp130 chain carrying the STAT3 recruitment site. Functional complementation of the STAT3 signaling pathways can only result in case of a molecular interaction between bait and prey and is measured via a luciferase reporter containing the rat pancreatitis-associated protein I (rPAP) promoter responsive to STAT3 [28]. A collection of 10,000 ORFs was screened for interactions with cytoplasmic domains of different PCDHs. Binary MAPPIT assays were performed to confirm the interactions and to select common candidate interactors. A shortlist of selected candidate interaction candidates has then been challenged in co-immunoprecipitation experiments.

III.2 Materials and Methods

III.2.1 Plasmids

LR-bait constructs were generated as previously described and named pCLG[27]. This constructs contains the bait of interest C-terminally fused to the long receptor isoform of the murine leptin receptor (LR), truncated after the JAK binding site and fused to a 20xGGS hinge. The pCLG vector is derived from pcDNA5/FRT plasmid (Flp-In system; Life technologies): it includes an FRT site related to the hygromycin resistance gene for Flp recombinase dependent integration to the host cell line, in this case TRex44.[29] The MAPPIT pCLG-hPCDHs bait vectors were generated by cloning the full cytoplasmic tail of the respective PCDHs or fragments in the pCLG vector backbone into the BamHI-XhoI or BamHI-NotI sites, as described in [30]. MAPPIT preys were either cloned into the EcoRI-XhoI or EcoRI-XbaI site of the pMG2 plasmid or were obtained by Gateway recombinationatorial transfer of the full size ORF cDNA of the candidates from an entry clone selected from the ORFeome collection version 5.1[31] into a pMG1 destination vector as reported before[32]. The generation of the STAT3-responsive luciferase reporter plasmid (pXP2d2-rPAP-luci) has been described before[33]. The reporter contains a STAT3-dependent promoter fragment derived from the rat Pancreatitis Associated Protein 1 (Pap) gene fused to luciferase.

III.2.2 Transient MAPPIT experiment

Hek293T cells were cultured in Dulbecco's modified Eagle's medium DMEM+10% fetal bovine serum, grown in a humidified atmosphere at 37°C and 5% CO_2, seeded in 96 microtiter well plates and transfected by calcium phosphate precipitation as previously described[33]. (Transfection reagents: 2.5 M $CaCl_2$. Prepared in distilled water, filter-sterilized by passage through a 0.45 µM nitrocellulose membrane and stored at 20°C. Hepes-Buffered Saline (HeBS): 280 mM NaCl, 1.5 mMNa_2HPO_4, 50 mM HEPES. Adjusted to pH to 7.05 with NaOH. Filter-sterilized by passage through a 0.45 µM nitrocellulose membrane and stored at 20°C. Briefly, for a transient MAPPIT experiment, cells were transfected after 24h from seeding in the 96 well plates with bait (25ng per well), prey (25ng per well) and reporter (5ng per well) plasmids. Samples were prepared in triplicates. 24h after transfections, triplicate wells were stimulated either with vehicle (media) or leptin (100ng/ml). Luciferase activity was measured using the Luciferase Assay System kit (Promega) on an Envision plate reader (Perkin Elmer) 24H after stimulation: cells were lysed for 10 minutes in 50 µl luciferase lysis buffer (25 mM Tris-phosphate, pH 7.8, 2 mM DTT, 2 mM 2,2

diaminocyclohexane-N,N,N0,N0-tetra-acetate (DCTA), 10 % glycerol, 1 % Triton X-100. Store at 20°C) and 35µl of Luciferase substrate buffer (40 mM tricine, 2.14 mM (MgCO3)4 Mg(OH)2 5H2O, 5.34 mM MgSO4 , 66.6 mM DTT, 0.2 mM EDTA, Coenzyme A, 734 µM Adenosine 50 triphosphate, 940 µM D-luciferin. Stored at 20°C). The reagent is light sensitive and 35 µl of luciferase substrate buffer were added just before measuring. Luciferase fold change was determined by calculating the ratio of leptin stimulated to unstimulated wells.

III.2.3 Array MAPPIT

Array MAPPIT screens were performed for PCDH10, PCDH11X and PCDH9 to identify novel common or exclusive interaction partners. The screens were performed using a subset of about 10.000 ORF preys selected from the human ORFeome collection version 5.1[31] as previously described[34]. Briefly, Hek293T cells were transfected in buck with the pCLG-PCDH9, -10 or -11X bait encoding plasmid; cells were plated 24h after transfection in array screening plates already printed with the prey and reporter reverse transfection mixtures. 24 hours later, each condition (prepared in duplicates) was stimulated with leptin (100ng/ml) and 24 hours later luciferase activity was detected and measured. The ratio between the average values of the two leptin stimulated samples were calculated and given as the MAPPIT signal. The signal intensity ranking list of preys was created for each screen. Preys were ranked also according to within experiment variation applying a one-sided balanced test[35].

III.3 RESULTS: MAPPIT technology

III.3.1 cDNA library MAPPIT experiments with PCDH cytoplasmic domains

The first MAPPIT screens, with the cytosolic domain of the δ-Protocadherin family, date back several years, when Dr. Uta Brunner from our group, performed a MAPPIT cDNA library screening for PCDH 11X/Y interaction partners. PCDH11 belongs to the δ1PCDH subfamily and it has been linked to tumor development. MAPPIT was picked as a promising approach not only because it allows the detection of protein interactions in mammalian cells but also because the interactions occur in proximity of the plasma membrane, the natural environment for PCDHs. In previous studies, Protein Phosphatase 1α (PP1α) has been identified, via a pull-down assay, as an interactor with the CM3 domain of δ1PCDHs including PCDH11[25]. PP1α plays a crucial role in the pRB tumor suppressor pathway, is essential for cell division, and participates in the regulation of glycogen metabolism, muscle contractility and protein synthesis but the role of its interaction with δ1PCDHs is still not clear[36].

Initial tests with the cytoplasmic tail of PCDH11X as bait and the known interactor PP1α as prey resulted in a clear-cut interaction, showing that MAPPIT can be used for this particular aim. In the same test, PCDH11X/Y and PCDH9 interacted with PP1α. The interaction screening was performed in a cell line derived from the Hek293 FlpIn T-Rex cell line (Invitrogen), the TRex44 MAPPIT screening cell line: this cell line allows the generation of a pool of isogenic cells expressing the chimeric bait receptor and the mouse virus receptor (mEcoR) to permit viral delivery of the prey cDNA library made as described in Lievens *et al.*[29]. These cells contain a reporter cassette encoding a STAT3-responsive rPAP1 promoter-driven hIL5Rα-derived membrane tag [29,37]. Retroviral transduction of the prey cDNA library into the bait-expressing cell pool is the first step of the screening protocol. The stimulation with ligand allows the selection of cells in which bait-prey interactions occur. Positive cells are enriched using magneto-beads thanks to the hIL5Rα membrane tag. By an additional cell sorting step the number of false positives can be reduced at this stage of the protocol. After a second stimulation with the ligand, single positive cells are spotted into multi-well plates. A dot blot assay is used for validation, and finally RT-PCR and sequencing is applied to reveal the identity of the preys.

This screening for PCDH11 interactors yielded a list of putative candidates, but none of them could be confirmed in a one-to-one, or binary MAPPIT re-test.

Several putative interactions partners (DYNLT1, FHL2, PDLIM7, FASCIN1, and Zyxin), suggested by the literature or by previous experiments, were subsequently cloned or obtained from the ORF prey collection to perform an analytical test using the full-length cytoplasmic tail of PCDH11X. This confirmed the interactions with the candidates, as expected. A clear interaction was identified between PCDH11X and DYNLT1. Interactions were mapped using different constructs for PCDH11X including all, some or any conserved motifs[25] (Uta Brunner, data not published).

A MAPPIT cDNA library screen was also performed for the identification of interactors of the PCDH10 cytoplasmic domain (Irene Kahr, data not published). At the time of the experiment no interactor of PCDH10 was known, so the interaction between PCDH11 or PCDH9 and PP1α was used as positive control and the bait was tested for interaction–independent activation via transfection with and without an irrelevant prey, in this case SV40 large-T antigen (SVT). The screen was performed in the same cell line and with the same protocol as in the case of the previous test on PCDH11.

Also in this case, as for the abovementioned PCDH11 cDNA library screening, the results were disappointing: sequencing of PCR products and consecutive BLAST searches showed only 14 readable sequences, and none of these were in frame with gp130. The MAPPIT cDNA screen was therefore not repeated: auto activation for the PCDH10 bait can lead to spontaneously hIL5RαΔcyt surface tag expression at a level prohibiting selection of true PPIs. Despite this failure of the MAPPIT cDNA library screen, Y2H screens were performed but also these failed in identifying putative interaction partners of PCDH10. Nonetheless, a binary MAPPIT experiment was performed using the PCDH10 cytoplasmic domain as bait: putative interaction partners of PCDH11 identified in our laboratory by Uta Brunner via MAPPIT or Y2H cDNA library screens (data not published) were retested in such one-to-one experiment. The results for PCDH10 demonstrated a clear interaction with one of the putative interaction partners of PCDH11, DYNLT1, also known as Tctex-1 (PhD project of Irene Kahr, data not published). DYNLT1 was then used as positive control in all the following MAPPIT experiments.

III.3.2 Array MAPPIT

Instead of repeating the experiment using the FACS based MAPPIT screening methods, we applied the newly developed Array MAPPIT[37] technology (for further details we refer to Chapter II) to screen for novel interaction partners. As baits we used the cytoplasmic domain of human PCDH9, PCDH10 and PCDH11X, as preys the library built on the ORFeome v5.1. The screens were performed in Hek293T cells and in an automated 384-well plate format. By comparing the results of the screening of different PCDHs and evaluating the relevance of the putative interaction partners for our research, we could establish a priority list of interaction candidates (**Figure 47**).

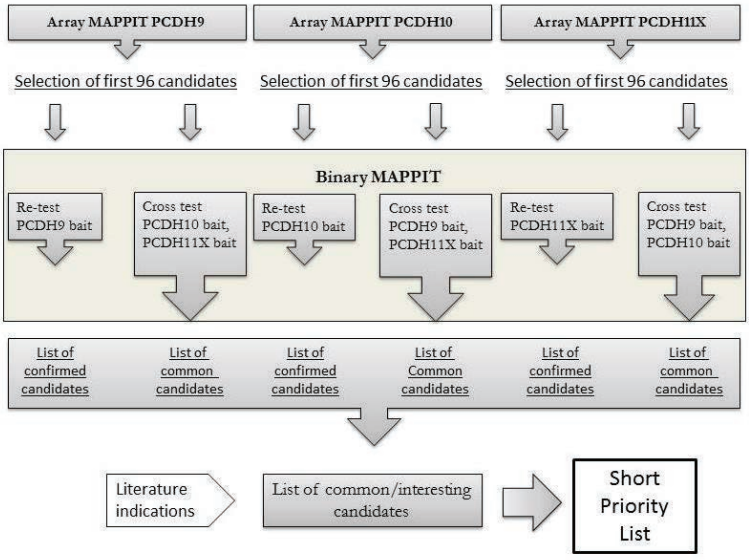


Figure 47 Summary of the MAPPIT procedure presented in this chapter and used to obtain the short priority list of common/exclusive candidate interaction partners of different members of the δPCDH family

The test is based on the activation of a luciferase reporter gene and the duration of the analysis process is dramatically reduced (5 days from transfection to measurement) since the preys can be identified simply by their position on the array. The method takes advantage of the availability of a large collection of the full-length human ORFs which can be transferred easily in the prey vector by recombinational cloning[38].

The general protocol of the array is based on a reverse transfection. First, microtiter plates containing plasmid preys covering the complete prey collection need to be generated. Then the screening proceeds with the covering of the plates by the bait-expressing cell pool, the stimulation with the appropriate ligand and the measurement of the luciferase activity. This approach is really fast and mostly automated[39]. For the experiments with PCDHs we did not use a stably transfected cell line but we proceeded with transiently transfected cells using the calcium phosphate protocol. Currently, a collection of 15,000 preys, corresponding to the ORFeome v8.1 collection (Centre for Cancer Systems Biology, CCSB, Boston), is available for screening. At the time of the experiment the collection included 10,000 preys. The absolute counts (triplicate for each condition, stimulated and non-stimulated) are analyzed and used to calculate the fold induction for each prey: this will be compared with the fold induction value of the empty prey which is used as threshold. The results come ranked based on fold induction and are represented on a plot. For each Array MAPPIT screen, a list with the top 96 hits is retested in a binary MAPPIT set up to check for specificity.

II.3.2.1 Array MAPPIT analysis for the PCDH11X cytoplasmic domain

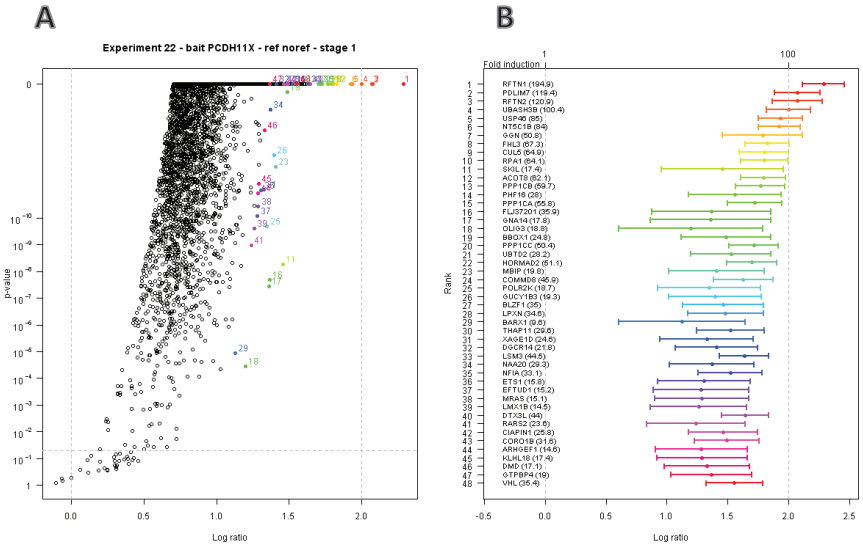


Figure 48 Array MAPPIT screen for PCDH11X

A) Every condition was screened in duplicates. The volcano plot represents the logarithmic (log) scale (x axis) of the ratio (fold change) of the leptin stimulated condition to the unstimulated condition and the p-value calculated using one sided balanced test. At the top right the most prominent candidates are gathered combining the high MAPPIT signal with high statistical significance. B) The top ranked 48 candidates.

The Array MAPPIT experiment for the PCDH11X cytoplasmic domain was performed and results are shown in **Figure 48**.

At first glance, we noticed that PP1α appears, as expected, among the top hits (PPP1CA or Serine/threonine-protein phosphatase PP1-alpha catalytic subunit, position 15 in the ranking list), confirming the validity of the test. Interestingly, also the catalytic subunits of PP1β (i.e. PPP1CB) and PP1γ (i.e. PPP1CC) are among this top list (position 13 and 20, respectively), indicating that δ1PCDHs might interact directly or indirectly with each of the PP1 isoforms.

The second protein on the list (PDLIM7, or Enigma) was already detected in the short list of candidates specifically tested in a one-to-one experiment. It is a member of a family of proteins that contain N-terminal PDZ domains and C-terminal LIM domains and probably functions as an adapter. It is known to associate with actin filaments of skeletal muscle. However, it also localizes via its PDZ domain to the actin microfilament network of non-muscle cells[40], indicating that its function extends to other tissues. Microscopic analysis showed co-localization of PDLIM7 with PCDH11X (Uta Brunner, unpublished data).

Next we proceeded to the specificity test, performing binary assays on the top 96 candidates (**Figure 49**). In the retest as a negative control we used an empty vector containing gp130 only and as positive control known JAK2 binding proteins such as REM2 GTPase and EF-hand domain-containing family member A1 (EFHA1). As internal control we also used two irrelevant baits: the DHFR from *E. coli* and bromodoman 4 C-terminal domain (Brd4-CTD). This control can exclude false positives due to the binding of the preys to the JAK2 molecules linked to the leptin receptor and can ensure that the interactors are specific for the selected PCDH.

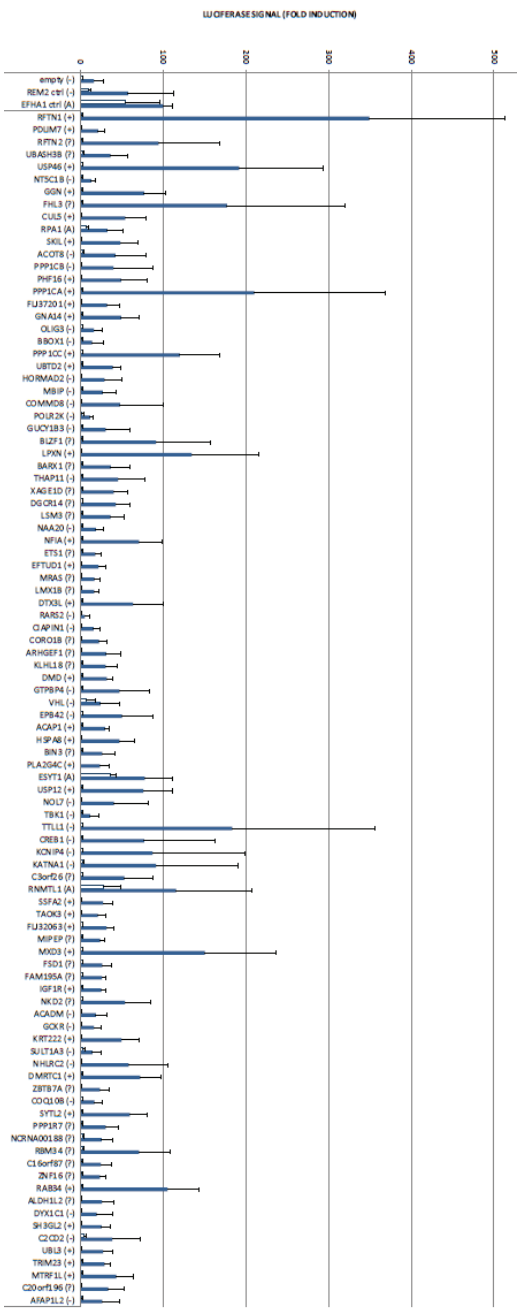


Figure 49 Re-test for candidate interactors of the PCDH11X cytoplasmic domain in a one-to-one MAPPIT assay

We used the full length cytoplasmic tail of PCDH11X; the test confirmed known and putative interaction partners, demonstrating the validity of the Array MAPPIT test. White: non-relevant eDHFR bait, Blue: original PCDH11X bait.

The experiment was performed twice with comparable results allowing the selection of a short list of candidates. On the basis of literature indications, also a few candidates, which turned positive in only one of the two tests, were retained. 29 interaction partners were confirmed for PCDH11X and these are listed in **Table 8.** As mentioned above, δPCDHs share several domains (CMs) in the cytoplasmic tail. To identify common interaction partners we decided to perform a cross test using the cytosolic domain of PCDH9, PCDH10 and PCDH11X as baits (details and explanation in the next paragraph). The cross test re confirmed 25 of the 29 partners. The complete list was subjected to an Ingenuity Pathway analysis in order to identify common associated pathways (**Figure 50**).

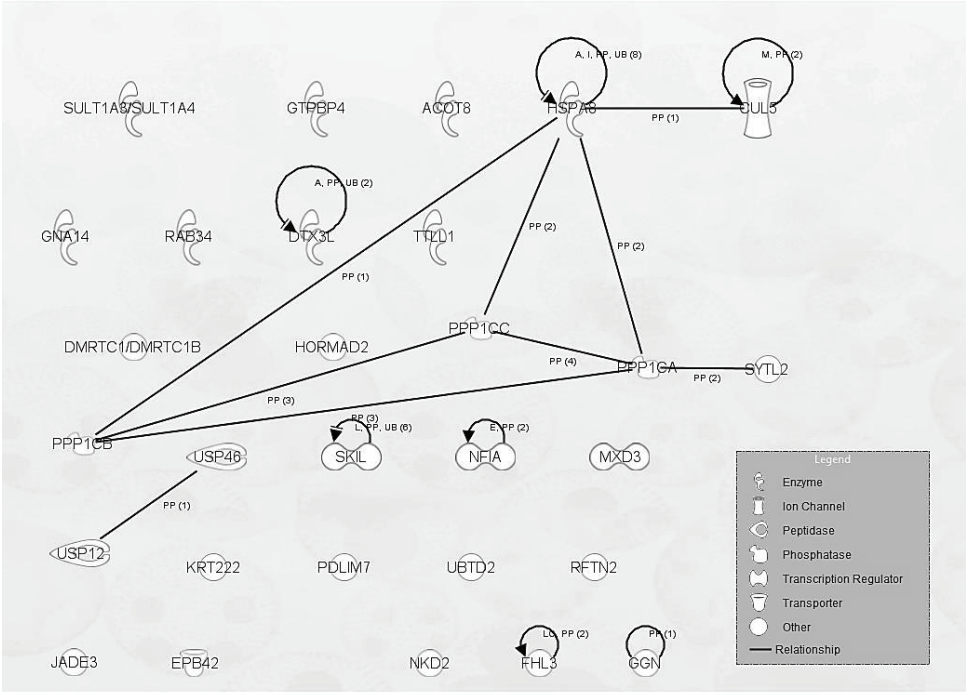


Figure 50 Analysis of interactions between confirmed candidates of PCDH11X with Ingenuity Pathway System.

As expected the system put in relationship PPP1CA, PPP1CB and PPP1CC, but also HSPA8, heat shock 70kDa protein 8, with CUL5.

Table 8 List of positive candidates for PCDH11X after MAPPIT experiments. Table obtained via IPA.

	Symbol	Entrez Gene Name	Location
1	ACOT8	acyl-CoA thioesterase 8	Cytoplasm
2	CUL5	Cullin 5	Nucleus
3	DMRTC1	DMRT-like family C1B	Other
4	DTX3L	deltex 3 like, E3 ubiquitin ligase	Cytoplasm
5	EPB42	erythrocyte membrane protein band 4.2	Plasma Membrane
6	FHL3	four and a half LIM domains 3	Plasma Membrane
7	GGN	Gametogenetin	Nucleus
8	GNA14	guanine nucleotide binding protein (G protein), alpha 14	Plasma Membrane
9	GTPBP4	GTP binding protein 4	Nucleus
10	HORMAD2	HORMA domain containing 2	Nucleus
11	HSPA8	heat shock 70kDa protein 8	Cytoplasm
12	KRT222	keratin 222, type II	Other
13	MXD3	MAX dimerization protein 3	Nucleus
14	NFIA	nuclear factor I/A	Nucleus
15	NKD2	naked cuticle homolog 2 (Drosophila)	Nucleus
16	PDLIM7	PDZ and LIM domain 7 (enigma)	Cytoplasm
17	PPP1CA	protein phosphatase 1, catalytic subunit, alpha isozyme	Cytoplasm
18	PPP1CB	protein phosphatase 1, catalytic subunit, beta isozyme	Cytoplasm
19	PPP1CC	protein phosphatase 1, catalytic subunit, gamma isozyme	Nucleus
20	RAB34	RAB34, member RAS oncogene family	Cytoplasm
21	RFTN2	raftlin family member 2	Other
22	SKIL	SKI-like proto-oncogene	Nucleus
23	SULT1A3/SULT1A4	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3	Cytoplasm
24	SYTL2	synaptotagmin-like 2	Cytoplasm
25	TRIM23	Tripartite motif containing 23	Nucleus
26	TTLL1	tubulin tyrosine ligase-like family member 1	Extracellular Space
27	UBTD2	ubiquitin domain containing 2	Other
28	USP12	ubiquitin specific peptidase 12	Cytoplasm
29	USP46	ubiquitin specific peptidase 46	Other

II.3.2.2 Array MAPPIT analysis for the PCDH10 cytoplasmic domain

Of the δ2PCDHs, we decided to screen PCDH10 as bait using the 10,000 ORF prey library. The human PCDH10 gene is frequently silenced in several carcinomas, and its ectopic expression strongly suppresses tumor cell growth, migration and invasion[20,41–45]. Despite its obvious relevance to cancer, not much is known about PCDH10. Our Array MAPPIT screen produced a list of putative interaction partners. As expected for a δ2PCDH lacking the consensus PP1α-interaction motif CM3, PPP1CA is not included in this list. The results are shown in the volcano plot of **Figure 51** and the list of the top 96 candidates can be found in **Table 11.**

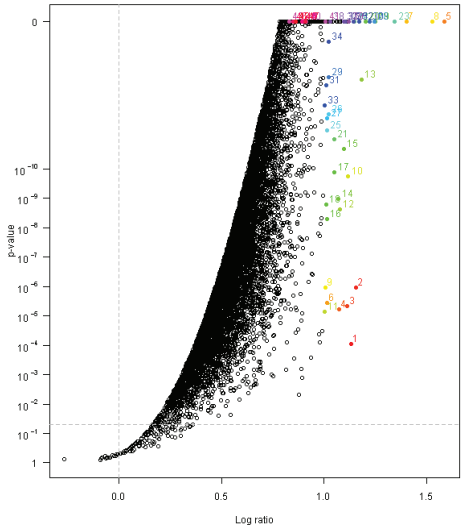


Figure 51 Array MAPPIT screen for PCDH10
Every condition was screened in duplicates. The volcano plot represents the logarithmic (log) scale (x axis) of the ratio (fold change) of the leptin-stimulated condition to the unstimulated condition and the p-value calculated using one sided balanced test. At the top right the most prominent candidates are gathered combining the high MAPPIT signal with high statistical significance.

We performed a re-test experiment for the top 96 candidates, to validate the results obtained from the Array MAPPIT. The analysis of this re-test was not as straightforward as might be expected for MAPPIT experiments under optimal conditions. This was due to a high background for the PCDH10 bait vector in our experiments. We obtained a final list of 25 candidate interactors. An experiment of "cross test" followed (see below). Of the 25 retained candidate interactors of the PCDH10 cytoplasmic domain, 12 turned out to be positive again in the cross test when PCDH10 was considered as bait (**Table 9**).

Table 9 List of positive candidates for PCDH10 after MAPPIT experiments. Table obtained via IPA.

	Symbol	Entrez Gene Name	Location
1	DMRTC1	DMRT-like family C1	Other
2	HEY1	hes-related family bHLH transcription factor with YRPW motif 1	Nucleus
3	HOXB7	homeobox B7	Nucleus
4	MXD3	MAX Dimerization Protein 3	Nucleus
5	MXI1	MAX interactor 1, dimerization protein	Nucleus
6	PDE1B	phosphodiesterase 1B, calmodulin-dependent	Cytoplasm
7	RGS20	regulator of G-protein signalling 20	Cytoplasm
8	RNF115	Ring Finger Protein 115	Cytoplasm
9	RRAGC	Ras-related GTP binding C	Cytoplasm
10	STAU1	staufen double-stranded RNA binding protein 1	Cytoplasm
11	TRIM23	tripartite motif containing 23	Nucleus
12	VGLL4	vestigial-like family member 4	Nucleus

II.3.2.3 Array MAPPIT analysis for the PCDH9 cytoplasmic domain

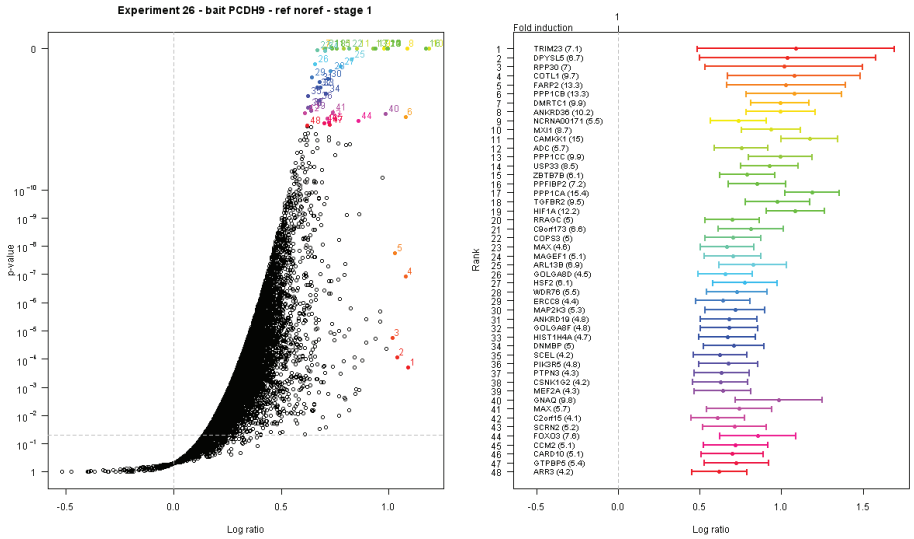


Figure 52 Array MAPPIT for PCDH9

Every condition was screened in duplicates. The volcano plot represents the logarithmic (log) scale (x axis) of the ratio (fold change) of the leptin stimulated condition to the unstimulated condition and the p-value calculated using one sided balanced test. At the top right of the volcano plot the most prominent candidates are gathered combining the high MAPPIT signal with high statistical significance. On the left of the figure, the top ranked 48 candidates

The Array MAPPIT screen using the PCDH9 cytoplasmic domain was also carried out using the same protocol as explained above. Results are shown in **Figure 52**. The list of the top 96 candidates has been validated via a re-test experiment.

The background signal for PCDH9 was significantly lower compared to this for other δPcdhs, and this allowed an easier selection. We obtained a list of 34 candidates (**Table 10**). 19 of them were confirmed in the subsequent cross-test.

Table 10 List of positive candidates for PCDH9 after MAPPIT experiments. Table obtained via IPA.

	Gene	Description	Location
1	AGBL3	ATP/GTP Binding Protein-Like 3	Cytoplasm
2	DPYSL5	dihydropyrimidinase-like 5	Cytoplasm
3	COTL1	coactosin-like F-actin binding protein 1	Cytoplasm
4	FARP2	FERM, RhoGEF and pleckstrin domain protein 2	Cytoplasm
5	PPP1CB	protein phosphatase 1, catalytic subunit, beta isozyme	Cytoplasm
6	PPP1CA	protein phosphatase 1, catalytic subunit, alpha isozyme	Cytoplasm
7	CAMKK1	calcium/calmodulin-dependent protein kinase 1, alpha	Cytoplasm
8	PPP1CC	protein phosphatase 1, catalytic subunit, gamma isozyme	Nucleus
9	DMRTC1	DMRT-like family C1	Other
10	TGFBR2	transforming growth factor, beta receptor II (70/80kDa)	Plasma Membrane
11	MXI1	MAX interactor 1, dimerization protein	Nucleus
12	USP33	ubiquitin specific peptidase 33	Cytoplasm
13	PPFIBP2	PTPRF interacting protein, binding protein 2 (liprin beta 2)	Nucleus
14	COPS3	COP9 signalosome subunit 3	Cytoplasm
15	GOLGA8DP	golgin A8 family, member D, pseudogene	Other
16	HSF2	heat shock transcription factor 2	Nucleus
17	SCEL	Sciellin	Cytoplasm
18	MEF2A	myocyte enhancer factor 2A	Nucleus
19	MAX	MYC associated factor X	Nucleus
20	C1orf106	chromosome 1 open reading frame 106	Other
21	HDAC11	histone deacetylase 11	Nucleus
22	FMN1	formin 1	Nucleus
23	CWC25	CWC25 spliceosome-associated protein homolog (S. cerevisiae)	Other
24	AGBL3	ATP/GTP binding protein-like 3	Other
25	B4GALT2	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 2	Cytoplasm
26	TRIM23	Tripartite motif containing 23	Nucleus
27	TRIT1	tRNA isopentenyltransferase 1	Cytoplasm
28	SETD2	SET domain containing 2	Cytoplasm
29	ZNF341	zinc finger protein 341	Nucleus

30	KLF17	Kruppel-like factor 17	Nucleus
31	SH3D19	SH3 domain containing 19	Plasma Membrane
32	DCAF12L1	DDB1 and CUL4 associated factor 12-like 1	Extracellular Space
33	PAX8	paired box 8	Nucleus
34	KLF4	Kruppel-like factor 4 (gut)	Nucleus

III.3.3 List of candidates: cross-test

Cross-tests were used to identify common interaction partners among different members of the δPCDH family. Each confirmed prey of each PCDH bait was re-tested in a binary test with the two other PCDH baits.

In the first cross-test PCDH9, PCDH10 and PCDH11X baits were screened for interactions with the confirmed candidates of PCDH11X. Few shared candidates were identified in the test: for example the PCDH10 bait gave a positive signal for PDLIM7 and NFIA, while PCDH9 was shown to be a candidate interactor for CUL5 and SKIL. Cross-tests have been performed with the same protocol for the confirmed candidates of PCDH10 and PCDH9. Positive interactions were added to the list of candidates for each bait. Results for each prey can be found in **Table 11.**

Table 11 Overall summary of our MAPPIT analyses.

A list of 96 top candidate interactors was obtained after every Array MAPPIT experiment. The first column shows the complete list of (96x3=) 288 preys. "Rank R (1)" values refer to the Array results for each experiment. The list of preys is alphabetically ordered so different colors are present to indicate the original array (orange, Array MAPPIT for the PCDH11X bait; blue, Array MAPPIT for the PCDH10 bait; yellow, Array Mappit for the PCDH9 bait). In the "Retest" column confirmed interactions are indicated with (+). In the last three columns, results are shown of three cross test experiments performed on interactions confirmed for each bait. (+)= interaction; (-) = no interaction; (?)=result missing/not trustable. In green are indicated the proteins included in the definitive selection. The final priority list was based on this combinatorial selection.

Prey name	Rank R(1)	PCDH	Retest	Cross test PCDH9	Cross test PCDH10	Cross test PCDH11X
ACADM	73	PCDH11X				
ACAP1	50	PCDH11X				
ACOT8	12	PCDH11X	(+)	(-)	(-)	(+)
ADC	15	PCDH9				
ADH4	11	PCDH10				
AFAP1L2	96	PCDH11X				
AGBL3	64	PCDH9	(+)	(+)	(-)	(+)
AGK	75	PCDH9				
ALDH1L2	88	PCDH11X				
ALG11	92	PCDH10				
ANKRD19	31	PCDH9				
ANKRD36	19	PCDH9				
APOL4	19	PCDH10				
ARHGAP28	61	PCDH9				
ARHGDIG	21	PCDH10				
ARHGEF1	44	PCDH11X				
ARL13B	25	PCDH9				
ARL17A	44	PCDH10				
ARNTL	94	PCDH10				
ARR3	48	PCDH9				
ATG3	7	PCDH10				
ATP5J2	73	PCDH9				
B4GALT2	66	PCDH9	(+)	(+)	(-)	(+)
BAIAP2	37	PCDH10				
BAIAP3	88	PCDH9				
BARX1	29	PCDH11X				
BBOX1	19	PCDH11X				
BIN3	52	PCDH11X				
BLZF1	27	PCDH11X				
BRMS1	16	PCDH10				
C10orf104	68	PCDH10				
C10orf32	74	PCDH9				
C16orf87	85	PCDH11X				
C17orf101	75	PCDH10				
C18orf32	67	PCDH10				
C1orf106	50	PCDH9	(+)	(?)	(+)	(+)
C20orf196	95	PCDH11X				
C2CD2	91	PCDH11X				
C2orf15	42	PCDH9				
C3orf26	62	PCDH11X				
C5orf32	72	PCDH10				
C8orf31	49	PCDH9				
C9orf131	89	PCDH10				
C9orf173	21	PCDH9				
CAMKK1	16	PCDH9	(+)	(-)	(-)	(-)
CAPN7	57	PCDH10				
CARD10	46	PCDH9				
CCDC21	35	PCDH10				
CCDC94	30	PCDH10				
CCM2	45	PCDH9				
CDC45L	18	PCDH10				

CDKN1B	47	PCDH10				
CHRNA1	69	PCDH9				
CIAPIN1	42	PCDH11X				
CLIP3	36	PCDH10				
CNN1	57	PCDH9	(+)	(-)	(+)	(-)
COMMD8	24	PCDH11X				
COPS3	21	PCDH9	(+)	(?)	(-)	(+)
COQ10B	80	PCDH11X				
CORO1B	43	PCDH11X				
COTL1	4	PCDH9	(+)	(+)	(-)	(+)
CREB1	59	PCDH11X				
CREB5	26	PCDH10				
CSNK1G2	38	PCDH9				
CSRP2BP	23	PCDH10				
CUL5	9	PCDH11X	(+)	(+)	(-)	(+)
CWC25	56	PCDH9	(+)	(+)	(+)	(+)
CXorf1	55	PCDH9				
DCAF12L1	87	PCDH9	(+)	(?)	(+)	(+)
DCUN1D1	95	PCDH10				
DGCR14	32	PCDH11X				
DMD	46	PCDH11X				
DMRTC1	78	PCDH11X	(+)	(+)	(-)	(+)
DMRTC1	20	PCDH9	(+)	(?)	(+)	(?)
DMRTC1	1	PCDH10	(+)	(+)	(+)	(+)
DMRTC2	24	PCDH10	(+)	(-)	(-)	(-)
DNMBP	34	PCDH9				
DPYSL3	90	PCDH10				
DPYSL5	2	PCDH9	(+)	(+)	(-)	(-)
DSN1	85	PCDH9				
DTX3L	40	PCDH11X	(+)	(-)	(-)	(+)
DYX1C1	89	PCDH11X				
EFTUD1	37	PCDH11X				
EIF2S2	66	PCDH10				
EIF3I	73	PCDH10				
EME1	96	PCDH10				
EPB42	49	PCDH11X	(+)	(-)	(-)	(+)
ERCC8	29	PCDH9				
ERF	3	PCDH10				
ESYT1	54	PCDH11X				
ETS1	36	PCDH11X				
FAF2	46	PCDH10				
FAM195A	70	PCDH11X				
FARP2	5	PCDH9	(+)	(-)	(-)	(-)
FCGR3A	92	PCDH9				
FGD2	81	PCDH9				
FHL3	8	PCDH11X	(+)	(-)	(-)	(+)
FLJ32063	66	PCDH11X				
FLJ37201	16	PCDH11X				
FMN1	52	PCDH9	(+)	(+)	(+)	(+)
FOXO3	44	PCDH9				
FOXP4	77	PCDH10				
FSD1	69	PCDH11X				
GCKR	74	PCDH11X				
GGN	7	PCDH11X	(+)	(+)	(-)	(+)
GKAP1	93	PCDH9	(+)			
GNA14	17	PCDH11X	(+)	(-)	(-)	(-)
GNAQ	40	PCDH9				
GOLGA8D	26	PCDH9	(+)	(+)	(+)	(+)
GOLGA8E	29	PCDH10	(+)	(+)	(?)	(+)
GOLGA8F	33	PCDH9				
GTPBP4	47	PCDH11X	(+)	(-)	(-)	(+)
GTPBP5	47	PCDH9				
GTPBP8	58	PCDH10				
GUCY1B3	26	PCDH11X				
HDAC11	51	PCDH9	(+)	(?)	(+)	(+)

HEY1	59	PCDH10	(+)	(+)	(+)	(-)
HIF1A	9	PCDH9				
HIST1H4A	32	PCDH9				
HORMAD2	22	PCDH11X	(+)	(-)	(-)	(+)
HOXB7	12	PCDH10	(+)	(?)	(+)	(+)
HPRT1	91	PCDH9				
HSF2	27	PCDH9	(+)	(?)	(?)	(?)
HSPA8	51	PCDH11X	(+)	(-)	(-)	(+)
IGF1R	71	PCDH11X				
ING2	55	PCDH10				
KATNA1	61	PCDH11X				
KCNIP4	60	PCDH11X				
KCNIP4	13	PCDH10				
KLF17	80	PCDH9	(+)	(+)	(+)	(+)
KLF4	94	PCDH9	(+)	(+)	(+)	(?)
KLF6	51	PCDH10				
KLHL18	45	PCDH11X				
KRT222	75	PCDH11X	(+)	(-)	(-)	(?)
LIG1	60	PCDH10				
LMX1B	39	PCDH11X				
LOC283693	50	PCDH10				
LPXN	28	PCDH11X				
LSM3	33	PCDH11X				
LUC7L3	78	PCDH10				
MAGEF1	24	PCDH9				
MAP2K3	30	PCDH9				
MAP2K3	32	PCDH10				
MAX	23	PCDH9				
MAX	41	PCDH9	(+)	(+)	(+)	(+)
MAX	6	PCDH10	(+)	(+)	(-)	(+)
MBIP	23	PCDH11X				
MEF2A	39	PCDH9	(+)	(+)	(+)	(+)
MEF2A	40	PCDH10				
MIER1	86	PCDH10				
MIPEP	67	PCDH11X				
MKKS	62	PCDH10				
MRAS	38	PCDH11X				
MTRF1L	94	PCDH11X				
MXD3 (MAD3)	68	PCDH11X	(+)	(+)	(-)	(+)
MXD3 (MAD3)	1	PCDH10	(+)	(+)	(+)	(+)
MXD4	5	PCDH10				
MXI1	17	PCDH9	(+)	(?)	(-)	(+)
MXI1	22	PCDH10	(+)	(-)	(-)	(+)
NAA20	34	PCDH11X				
NCRNA00171	18	PCDH9				
NCRNA00188	83	PCDH11X				
NFIA	35	PCDH11X	(+)	(-)	(-)	(+)
NHLRC2	77	PCDH11X				
NKD2	72	PCDH11X	(+)	(-)	(-)	(+)
NLRP12	27	PCDH10				
NOL7	56	PCDH11X				
NT5C1B	6	PCDH11X				
OLIG3	18	PCDH11X				
PANK1	63	PCDH10				
PAX8	90	PCDH9	(+)	(+)	(-)	(+)
PBX2	58	PCDH9				
PDE1B	81	PCDH10	(+)	(+)	(+)	(-)
PDHA2	71	PCDH9				
PDIK1L	61	PCDH10				
PDLIM7	2	PCDH11X	(+)	(-)	(-)	(+)
PHF16	14	PCDH11X				

PIK3R5	36	PCDH9				
PITPNB	15	PCDH10				
PIWIL1	14	PCDH10	(+)(?)			
PLA2G4C	53	PCDH11X				
PLCXD1	93	PCDH10				
POLR2K	25	PCDH11X				
PPFIBP2	11	PCDH9	(+)	(+)	(+)	(-)
PPP1CA	15	PCDH11X	(+)	(+)	(+)	(+)
PPP1CA	10	PCDH9	(+)	(+)	(-)	(+)
PPP1CB	13	PCDH11X	(+)	(+)	(-)	(-)
PPP1CB	6	PCDH9	(+)	(+)	(-)	(+)
PPP1CC	20	PCDH11X	(+)	(+)	(-)	(+)
PPP1CC	6	PCDH9	(+)	(?)	(-)	(+)
PPP1R7	82	PCDH11X				
PRIM2	8	PCDH10				
PRKCZ	65	PCDH9				
PRKRA	76	PCDH10	(+)	(+)	(?)	(+)
PRKRIR	41	PCDH10				
PSMC1	39	PCDH10	(?)	(?)	(?)	(+)
PSPH	84	PCDH9				
PSTPIP1	54	PCDH9				
PTPN3	37	PCDH9				
PYCR1	64	PCDH10				
RAB34	87	PCDH11X	(+)	(-)	(?)	(+)
RAB7L1	56	PCDH10				
RAB9A	31	PCDH10				
RALYL	79	PCDH9				
RAP1GDS1	74	PCDH10	(+)	(+)	(?)	(-)
RARS2	41	PCDH11X				
RASGEF1B	67	PCDH9				
RASL11B	70	PCDH10				
RBM34	84	PCDH11X				
RFTN1	1	PCDH11X				
RFTN2	3	PCDH11X	(+)	(-)	(-)	(+)
RGS20	87	PCDH10	(+)	(-)	(+)	(-)
RLIM	79	PCDH10				
RNF113A	71	PCDH10				
RNF115	17	PCDH10	(+)	(-)	(+)	(-)
RNF20	84	PCDH10				
RNMTL1	63	PCDH11X				
RPA1	10	PCDH11X				
RPP30	3	PCDH9				
RRAGC	7	PCDH9				
RRAGC	85	PCDH10	(+)	(-)	(+)	(-)
SCEL	35	PCDH9	(+)	(+)	(-)	(+)
SCRN2	43	PCDH9				
SETD2	77	PCDH9				
SH3D19	83	PCDH9	(+)	(+)	(+)	(+)
SH3GL2	90	PCDH11X				
SKIL	11	PCDH11X	(+)	(-)	(-)	(+)
SNF8	65	PCDH10				
SORBS3	38	PCDH10				
SPDYE2	96	PCDH9				
SPDYE2	20	PCDH10	(+)	(-)	(-)	(+)
SSFA2	64	PCDH11X				
STAU1	82	PCDH10	(+)	(+)	(+)	(+)
SULT1A3	76	PCDH11X	(+)	(-)	(-)	(+)
SYT14	95	PCDH9				
SYTL2	81	PCDH11X	(+)	(-)	(-)	(+)
TAOK3	65	PCDH11X				
TAS2R46	63	PCDH9				
TBC1D23	72	PCDH9				
TBCEL	28	PCDH10				
TBK1	57	PCDH11X				
TBRG1	53	PCDH9				

TBX20	76	PCDH9				
TGDS	48	PCDH10				
TGFBR2	9	PCDH9	(+)	(-)	(-)	(-)
THAP11	30	PCDH11X				
TNNI2	91	PCDH10				
TP53I11	86	PCDH9				
TPP2	80	PCDH10				
TRIM23	93	PCDH11X	(+)	(-)	(?)	(-)
TRIM23	1	PCDH9	(+)	(+)	(+)	(+)
TRIM23	33	PCDH10	(+)	(+)	(?)	(+)
TRIT1	68	PCDH9	(+)			
TROVE2	42	PCDH10				
TSC1	62	PCDH9				
TTLL1	58	PCDH11X	(+)	(-)	(-)	(+)
TUBGCP3	52	PCDH10				
TXNRD2	54	PCDH10				
U2AF1	53	PCDH10				
UBASH3B	4	PCDH11X				
UBC	59	PCDH9				
UBE2E2	69	PCDH10				
UBL3	92	PCDH11X				
UBTD2	21	PCDH11X	(+)	(-)	(?)	(+)
USP12	55	PCDH11X	(+)	(-)	(-)	(+)
USP21	60	PCDH9				
USP33	13	PCDH9	(+)	(+)	(-)	(-)
USP46	5	PCDH11X	(+)	(-)	(-)	(+)
VDR	88	PCDH10				
VGLL4	49	PCDH10	(+)	(-)	(+)	(+)
VHL	48	PCDH11X				
VPS45	25	PCDH10	(+)	(+)	(?)	(+)
WDR76	28	PCDH9				
XAGE1D	31	PCDH11X				
YME1L1	83	PCDH10	(+)	(-)	(-)	(+)
ZBTB7A	79	PCDH11X				
ZBTB7B	12	PCDH9	(+)	(-)	(-)	(-)
ZBTB7B	4	PCDH10	(+)	(+)	(-)	(+)
ZIC4	70	PCDH9				
ZNF16	86	PCDH11X				
ZNF202	9	PCDH10	(+)	(+)	(-)	(+)
ZNF296	43	PCDH10	(+)	(?)	(?)	(+)
ZNF341	78	PCDH9	(+)			
ZNF341	45	PCDH10				
ZNF548	89	PCDH9				
ZNF711	82	PCDH9	(+)	(-)	(-)	(-)

III.3.4 Analysis of the results

Array Mappit was performed for 3 different PCDHs and gave a list of 96 top candidates for each bait. All these candidates were re-tested in a binary MAPPIT experiment for confirmation. Numbers of confirmed candidates are listed in **Table 12**. The cross test was used for two purposes: as an extra re-test for the bait of reference ("Cross test" in the Table) and to identify extra partners for the other baits ("Cross test other Pcdhs" in the Table). Each PCDH received a small number of extra candidate interactors when analyzed in the cross test experiment, and these were added to the confirmed list obtained after the re-tests.

Table 12 Summary of positives in different test for the analyzed PCDHs

	Array MAPPIT	Re test	Cross test	Cross test other Pcdhs
PCDH11X	96	29	25	38
PCDH10	96	25	12	15
PCDH9	96	34	19	21

Finally, we selected a priority list of candidate interactors. The challenge was then to prioritize putative interaction partners to be confirmed by biochemical methods. To this end deep literature search and IPA analysis were performed to find links and connections between candidates. At the time of this analysis we were particularly focused on the roles that different PCDHs could play in cancer. Different shortlists have been created. **Table 13** shows for each PCDH cytoplasmic domain the definitive selection of candidate interactors .This selection takes into account: the original array-MAPPIT, the re-tests, the cross-interactions and the literature selection.

Table 13 Common candidates, interesting partners on the basis of literature indications and interactors already identified in previous experiments have been selected and listed

	PCDH9	PCDH10	PCDH11X
CUL5	+	-	+
DMRTC1	+	+	+
FHL3	-	-	+
KNIP4	-	+	-
GGN	-	-	+
MAX	+	+	+
MXD3 (MAD3)	+	-	+
MXD4	+	-	+
MXI1	+	+	-
PDLIM7	+	-	+
TRIM23	-	+	+
USP12	-	-	+
USP33	+	-	-
USP46	-	-	+
ZBTBT7	+	+	+

From this table we made a first priority list for biochemical studies including MAX, MXD3(MAD3) and MXI1, all three belonging to the same pathway, and also CUL5, FHL3, PDLIM7. On the basis of literature indications, we added in our following experiments also some other candidates from the MAPPIT list that were not included in the top scores. This selection comprised PSMC1, USP family members and actin-related proteins. Before proceeding with CoIP experiments, a confirmation of those latter candidates in a binary MAPPIT assay was warranted. For the USP family we obtained a short list of candidate interactors for each PCDH (**Table 14**). In this confirmation experiment, we used as bait constructs, either plasmids encoding full-size cytoplasmic domains of PCDH11X and PCDH10, or encoding progressive truncations of these cytoplasmic domains. For details see **Figure 53**. Results are shown in **Figure 54** and summarized in **Table 15** and **Table 16**.

Table 14 list of Ubiquitin related putative interaction partner

PCDH11X	PCDH10	PCDH9
UBL3	PSMC1	USP33
UBASH3B	U2AF1	MEF2A
UBTD2	UBE2E2	UBC
USP12		USP21
CUL5		CUL5

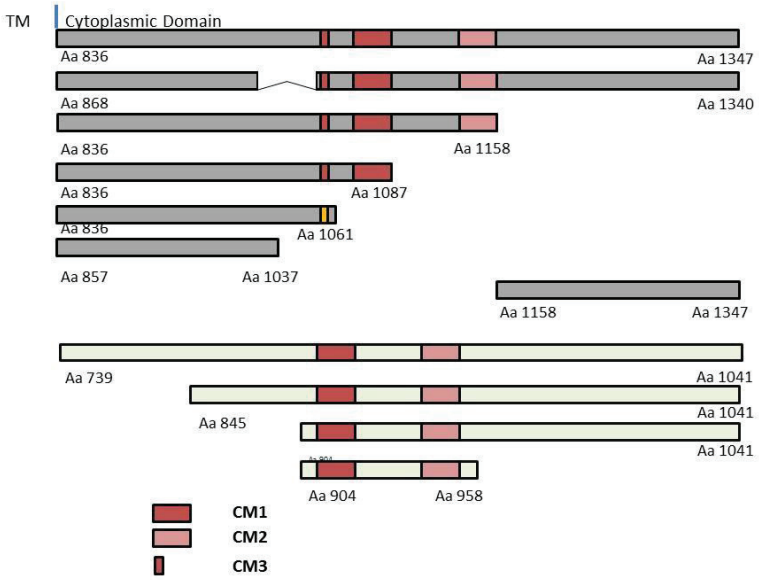
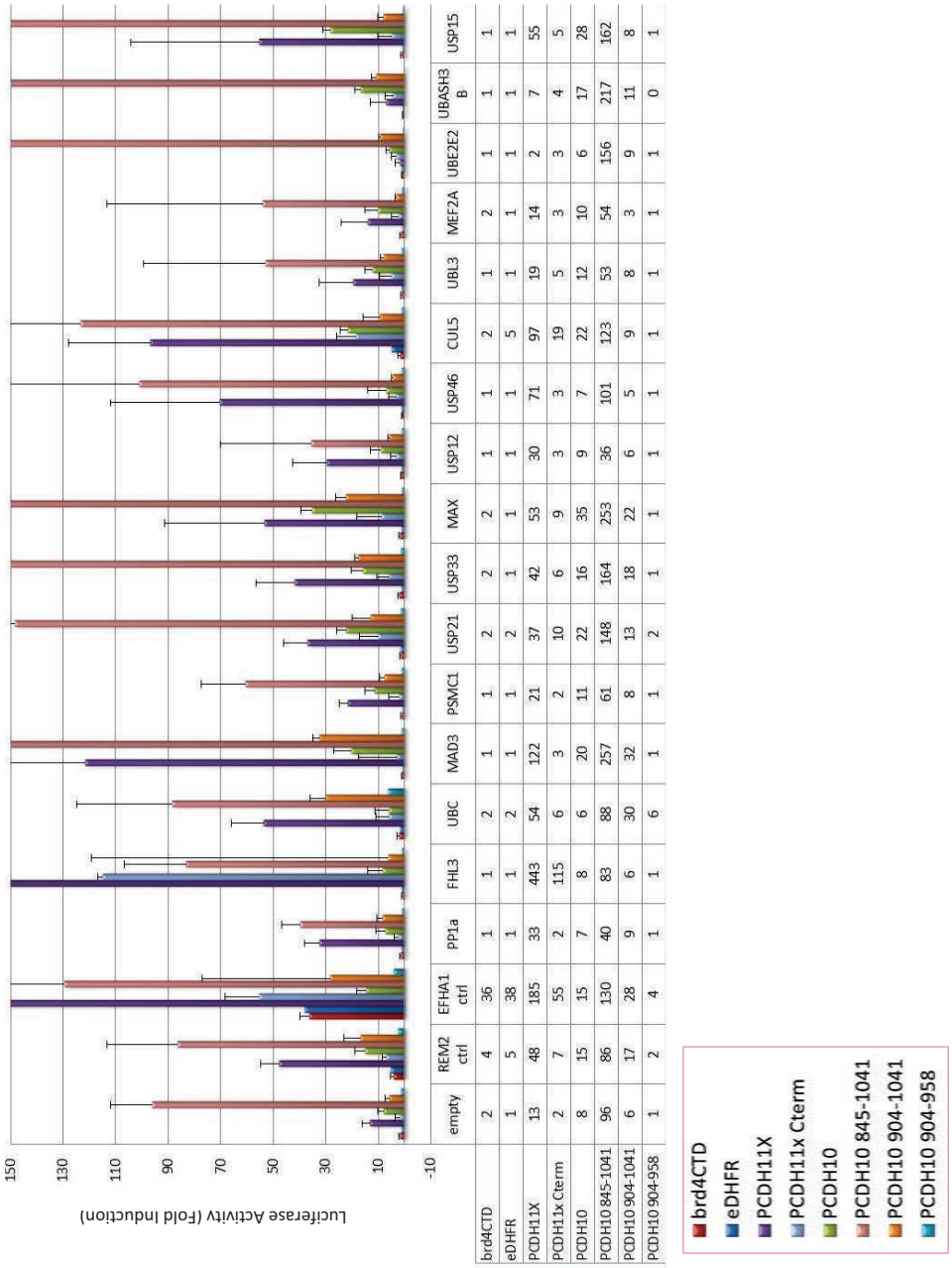


Figure 53 Full length and N-and C- terminally truncated constructs for the cytoplasmic domain of PCDH10 and PCDH11X.
Conserved motives are represented with colored boxes. CM, conserved motif; TM, transmembrane domain



	empty	REM2 ctrl	EFHA1 ctrl	PP1a	FHL3	UBC	MAD3	PSMC1	USP21	USP33	MAX	USP12	USP46	CUL5	UBL3	MEF2A	UBE2E2	UBASH3B	USP15
brd4CTD	2	4	36	1	1	2	1	1	2	2	2	1	1	2	1	2	1	1	1
eDHFR	1	5	38	1	1	2	1	1	2	1	1	1	1	5	1	1	1	1	1
PCDH11X	13	48	185	33	443	54	122	21	37	42	53	30	71	97	19	14	2	7	55
PCDH11x Cterm	2	7	55	2	115	6	3	2	10	6	9	3	3	19	5	3	3	4	5
PCDH10	8	15	15	7	8	6	20	11	22	16	35	9	7	22	12	10	6	17	28
PCDH10 845-1041	96	86	130	40	83	88	257	61	148	164	253	36	101	123	53	54	156	217	162
PCDH10 904-1041	6	17	28	9	6	30	32	8	13	18	22	6	5	9	8	3	9	11	8
PCDH10 904-958	1	2	4	1	1	6	1	1	2	1	1	1	1	1	1	1	1	0	1

Figure 54 One-to-one MAPPIT test for PCDH10cyt, PCDH11Xcyt and truncated vertions
Different constructs were screened with relevant preys such as USPs (and related) proteins, MAX and MXD3(MAD3).

Table 15 Confirmed (Grey) candidates for interaction with different constructs of PCDH11X and PCDH10 cytoplasmic domains

	USP21	USP46	CUL5	USP21	USP33	USP15	UBC	MAX	MAD3
PCDH11X									
PCDH11X C-terminal									
PCDH10									
PCDH10 845-1041									
PCDH10 904-1041									
PCDH10 904-958									

Table 15 only shows confirmed candidates for interaction with different constructs of PCDH11X and PCDH10 cytoplasmic domains. An additional binary MAPPIT experiment was finally performed with the same protocol in order to test for interaction with two additional PCDH11X fragment constructs and for additional candidate interactions of the priority list. These results are shown in **Table 16**.

Table 16 Confirmed (Grey) candidates for interaction with different constructs of PCDH11X and PCDH10 cytoplasmic domains

	MAX	MXD1	PDLIM7	MDM2	FHL3
PCDH11X					
PCDH11X CM3					
PCDH11X C-terminal					
PCDH11X CM3 MUT					

Based on literature suggestion, a binary MAPPIT assay was performed to specifically confirm the results obtained from the Array MAPPIT screens concerning actin related proteins. The one-to one experiment was performed for 7 different PCDH baits: PCDH9cyt, PCDH11Xcyt, PCDH11Xcyt T TERMINAL, PCDH11Ycyt, PCDH-PCcyt, PCDH18cyt and PCDH19cyt and 24 preys plasmids: 21 putative interaction partner, 1 empty prey and 2 positive controls (**Figure 55**).

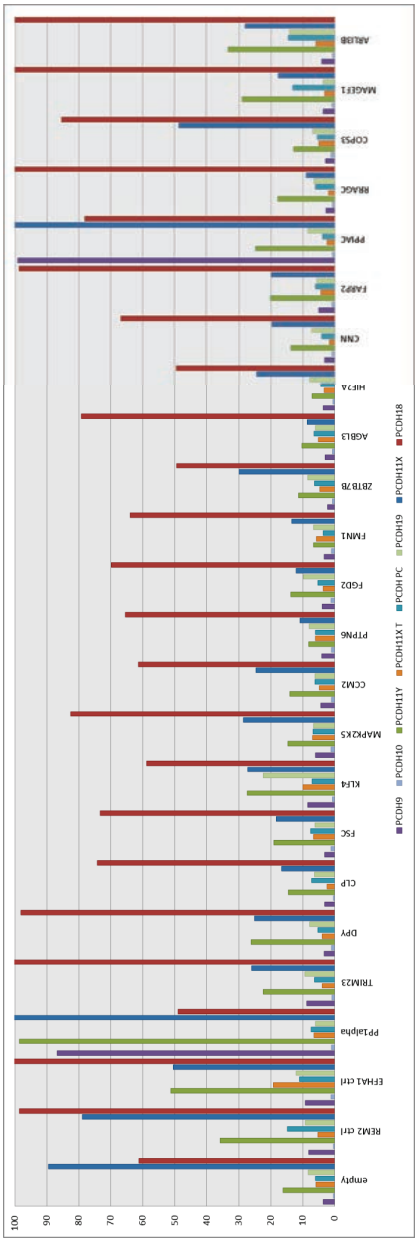


Figure 55 One-to-One MAPPIT test for actin related protein as preys and different PCDHs as baits.
On the left, list of preys. On the right the results. Y axis: Luciferase activity (fold induction)

From these results we could confirm some candidates suggested from the screens. In conclusion KLF4 is positive for PCDH9 and PCDH19, and ARL13B for PCDH11Y, PCDH18 and PCDH19. Kruppel-like factor 4 is a member of the KLF family of transcription factors and regulates proliferation, differentiation, apoptosis and somatic cell reprogramming.

II.3.5 Ingenuity Pathway Analysis

The Ingenuity Pathway Analysis (IPA) software has been developed to analyze and understand complex 'omics' data. Different kinds of data can be loaded and analyzed. We used the system to identify the relationship between the candidates of the obtained lists, checking for interactions but also for common functions. Each list was studied independently from the others concerning interactions and belonging to the same common pathway**. Figure 56, 57** and **58** show the involvement of different candidates from the MAPPIT lists in canonical relevant pathways.

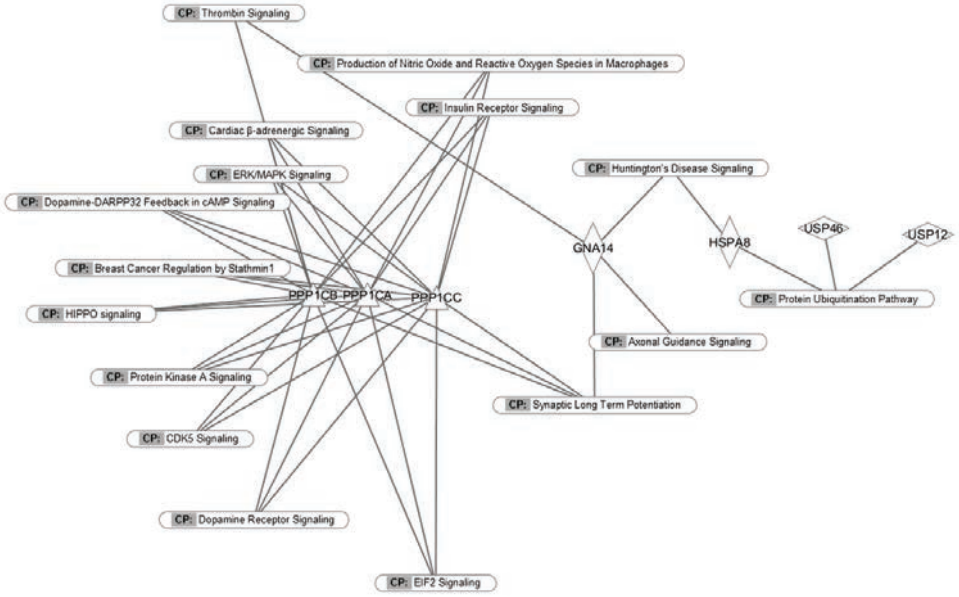


Figure 56 Canonical pathway related to PPP1CB,PPP1CA,PPP1CC, GNA14, HSP8 ans USP46 and USP 12

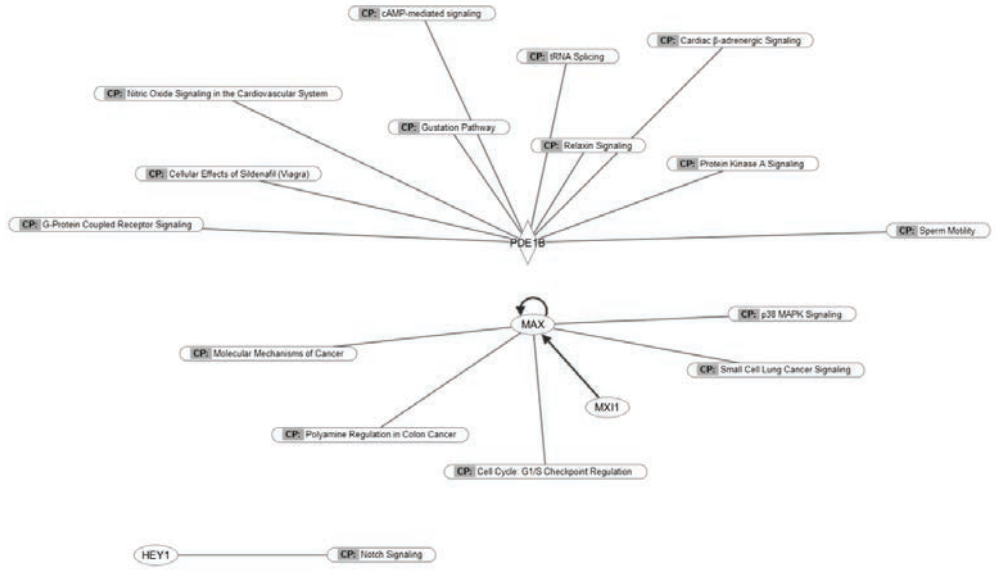


Figure 57 Canonical pathway related to MAX,MXI1, PDE18.

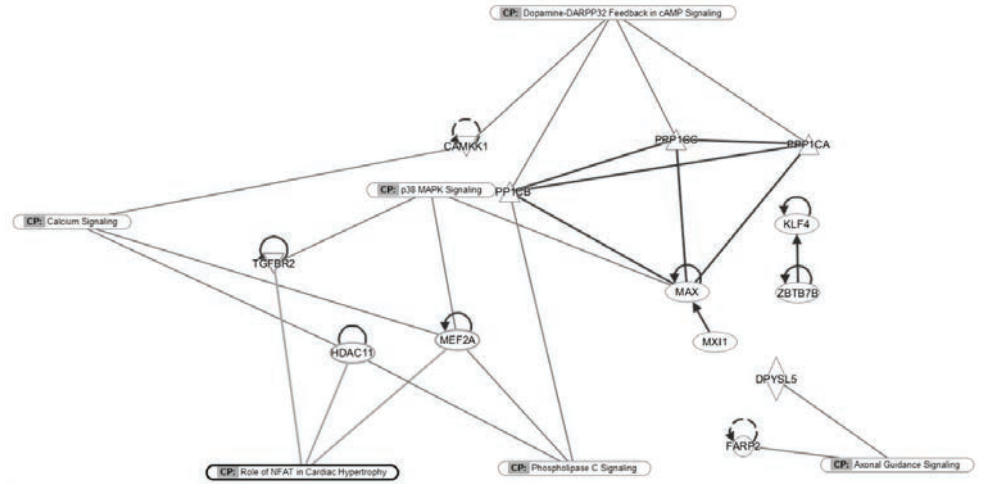


Figure 58 Canonical pathways related to TGFBR2, HDAC11, MEF2A, PP1CA, PP1CB, PP1CC, MAX and MXI1.

Next, the complete list of candidates was analyzed for already known interactions with PCDHs but no known link were identified. IPA has been used then to identify relationship between the family of PCDHs and the list of candidates including the shortest path which can connect two molecules of the

group. Only one node was added. The **Figure 59** shows the relationships. The nodes are listed in **Table 17**.

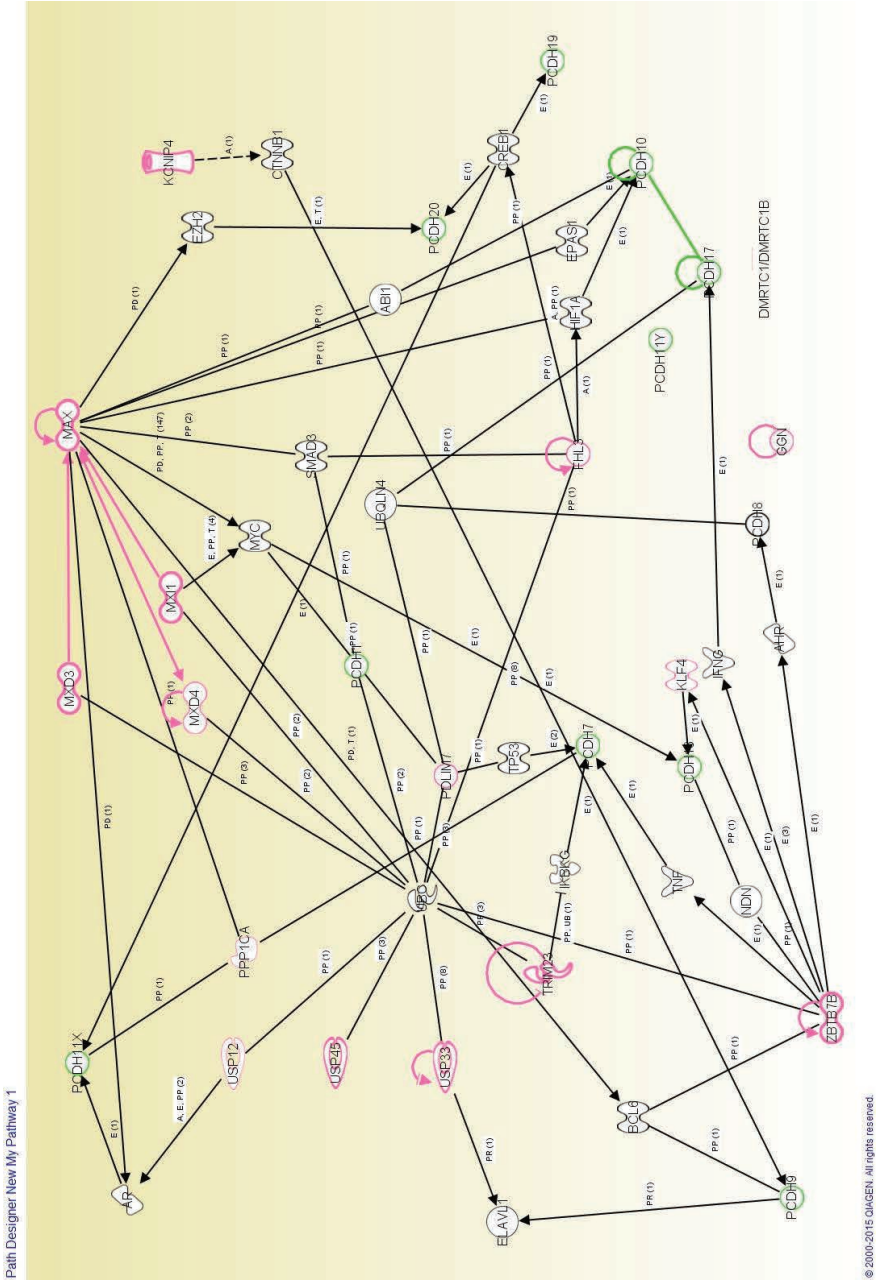


Figure 59 IPA Path Explorer. Candidates (pink) and PCDHs (green) relations in the shortest pathway

Table 17 PCDHs and candidates relations. List of protein of the shortest path. (IPA analysis)

Symbol	Synonym(s)	Location	Family
ABI1	Abelson interactor 1, Abl binding protein 4, abl-interactor 1, ABLBP4, E3B1, Eps8 binding, NAP1, Nap1 binding, NAP1BP, SSH3BP1, SSH3BP	Cytoplasm	Other
AHR	Ah, AH receptor, Ahh, Ahre, aryl-hydrocarbon receptor, bHLHe76, DIOXIN receptor, In	Nucleus	ligand-dependent nuclear receptor
AR	AIS, Andr, androgen receptor, AW320017, DHTR, HUMARA, HYSP1, KD, NR3C4, SBMA, SMAX1, Testosterone receptor, TFM	Nucleus	ligand-dependent nuclear receptor
BCL6	B cell leukaemia/lymphoma 6, B cell leukemia/lymphoma 6, B-cell CLL/lymphoma 6, BCL5, BCL6A, LAZ3, ZBTB27, zinc finger protein 51, ZNF51	Nucleus	transcription regulator
CREB1	2310001E10Rik, 3526402H21RIK, AV083133, CAMP RESPONSE ELEMENT binding protein 1, cAMP responsive element binding protein 1, Cbp, CREB	Nucleus	transcription regulator
CTNNB1	armadillo, Beta-cat, Beta-catenin, beta-Ctnn, Bfc, catenin (cadherin associated protein), beta 1, catenin (cadherin associated protein), β 1, catenin (cadherin-associated protein), beta 1, 88kDa, catenin (cadherin-associated protein), β 1, 88kDa, CATENIN beta, CATENIN β, CATNB, CTNN beta, CTNN β, CTNNB, Mesc, MRD19, β-cat, β-catenin, β-Ctnn	Nucleus	transcription regulator
DMRTC1/DMRTC1B	DMRT-like family C1, DMRT-like family C1B, DMRTC1, DMRTC1B	Other	Other
ELAVL1	2410055N02Rik, DKFZP667b083, ELAV1, ELAV (embryonic lethal, abnormal vision)-like 1 (Hu antigen R), ELAV like RNA binding protein 1, Hu antigen R, Hua, HUR, MelG, RGD:731215, RNA binding protein HuR, W91709	Cytoplasm	Other
EPAS1	bHLHe73, ECYT4, endothelial PAS domain protein 1, Hif2, HIF-2alpha, Hif1alpha related factor, HIF2 α, HIF2A, HLF, HRF, MOP2, PASD2	Nucleus	transcription regulator
EZH2	EHZ2, enhancer of zeste 2 polycomb repressive complex 2 subunit, enhancer of zeste homolog 2 (Drosophila), ENX-1, Enx1h, EZH1, EZH2b, KMT6, KMT6A, mKIAA4065, WVS2, WVS	Nucleus	transcription regulator
FHL3	four and a half LIM domains 3, SLIM2	Plasma Membrane	Other
GGN	AI593290, gametogenetin	Nucleus	Other
HIF1A	AA959795, bHLHe78, HIF1, HIF-1alpha (hydroxylated), HIF-1α, HIF-1α (hydroxylated), HIF1-ALPHA, Hypoxia inducible factor 1 alpha subunit, Hypoxia inducible factor 1 α subunit, hypoxia inducible factor 1, alpha subunit, hypoxia inducible factor 1, α subunit, hypoxia-inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor), hypoxia-inducible factor 1, α subunit (basic helix-loop-helix transcription factor), MOP1, PASD8	Nucleus	transcription regulator
IFNG	IFG, IFI, IFN gamma, IFN type II, IFN-γ, IFNG2, INF-γ, interferon gamma, Interferon γ, interferon, gamma, interferon, γ, type II INTERFERON, γ interferon, γ-ifn	Extracellular Space	Cytokine
IKBKG	1110037D23Rik, AI848108, AI851264, AMCBX1, AW124339, FIP-3, Fip3p, I Kappa B Gamma, I κ B γ, IkB γ,	Nucleus	Kinase

	IKK-gamma, IKK-γ, IKK[g], IKKAP1, IKKG, IMD33, inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase gamma, inhibitor of kappaB kinase gamma, inhibitor of kappaB kinase γ, inhibitor of κ light polypeptide gene enhancer in B-cells, kinase γ, IP1, IP2, IP, IPD2, NEMO, NEMO RELATED, ZC2HC9		
KCNIP4	AV032399, Calp250, CALP, KCHIP4, KChIP4a, Kv channel interacting protein 4	Plasma Membrane	ion channel
KLF4	EZF, GKLF, Kruppel-like factor 4 (gut), Zie	Nucleus	transcription regulator
MAX	AA960152, AI875693, bHLHd4, bHLHd5, bHLHd6, bHLHd7, bHLHd8, Max protein, MYC associated factor X, Myn	Nucleus	transcription regulator
MXD3	4631412E13Rik, BHLHC13, MAD3, MAX dimerization protein 3, MYX	Nucleus	transcription regulator
MXD4	2810410A03RIK, bHLHc12, MAD4, MAX dimerization protein 4, MST149, MSTP149	Nucleus	transcription regulator
MXI1	bHLHc11, ENSMUSG00000067085, Gm10197, LOC100360898, MAD2, Max inter 1, Max interacting protein 1, MAX interactor 1, MAX interactor 1, dimerization protein, MAXD2, MX11, MXD2, MXI, MXI-WR	Nucleus	transcription regulator
MYC	AU016757, bHLHe39, C-MYC-P64, CMYC, mMyc, MRTL, Myc2, MYCC, myelocytomatosis oncogene, Niard, Nird, RNCMYC, v-myc avian myelocytomatosis viral oncogene homolog	Nucleus	transcription regulator
NDN	AI528698, HsT16328, NECDIN, necdin, melanoma antigen (MAGE) family member, Peg6, PWCR	Nucleus	Other
PCDH1	2010005A06RIK, AI585920, AXIAL PROTOCADHERIN, Pc1, PC42, PCDH42, protocadherin 1	Plasma Membrane	Other
PCDH10	6430521D13RIK, 6430703F07Rik, mKIAA1400, OL-pc, OL-PCDH, PCDH19, protocadherin 10, RGD1565811	Plasma Membrane	Other
PCDH11X	A230092L07RIK, LOC100039461, LOC100046467, PCDH11, PCDH-X, PCDHX11, PPP1R119, protocadherin 11 X-linked, RGD1562864	Plasma Membrane	Other
PCDH11Y	PCDH22, PCDH-PC, PCDHX, PCDHY, protocadherin 11 Y-linked	Other	Other
PCDH17	C030033F14RIK, Gm78, LOC144997, PCDH68, PCH68, protocadherin 17	Plasma Membrane	Other
PCDH18	3110038E07Rik, BB095589, PCDH68L, protocadherin 18	Extracellular Space	Other
PCDH19	B530002L05Rik, EFMR, EIEE9, Gm717, LOC279653, mKIAA1313, protocadherin 19, RGD1565392	Extracellular Space	Other
PCDH20	C630015B17RIK, PCDH13, protocadherin 20	Other	Other
PCDH7	BH-Pc, BHPCDH, PPP1R120, protocadherin 7	Plasma Membrane	Other
PCDH8	1700080P15Rik, ARCADLIN, PAPC, protocadherin 8	Plasma Membrane	Other
PCDH9	C530050I23RIK, LOC638275, protocadherin 9	Plasma Membrane	Other
PDLIM7	1110003B01RIK, AV007930, Enigma, LMP1, LMP3, LMP, PDZ and LIM domain 7, PDZ and LIM domain 7 (enigma)	Cytoplasm	Other

PPP1CA	dism2, PP-1A, Pp-1c, PP1 α, PP1 α C, PP1alpha, PP1C alpha, PP1C α, PPP1A, Ppp1c, Protein Phosphatase 1, protein phosphatase 1, catalytic subunit, alpha isoform, protein phosphatase 1, catalytic subunit, alpha isozyme, protein phosphatase 1, catalytic subunit, α isoform, protein phosphatase 1, catalytic subunit, α isozyme, Ser/Thr Phosphatase Type1 alpha, Ser/Thr Phosphatase Type1 α	Cytoplasm	Phosphatase
SMAD3	AU022421, DKFZP586N0721, hMAD-3, HSPC193, HsT17436, JV15-2, LDS3, LDS1C, MAD3, MADH3, SMAD family member 3	Nucleus	transcription regulator
TNF	AT-TNF, DIF, RATTNF, TMTNF, TNF-a, TNF-alpha, TNF-α, TNFSF2, Tnfsf1a, tumor necrosis factor, Tumor Necrosis Factor α, tumor necrosis factor, α, tumour necrosis factor, tumour Necrosis Factor Alpha, tumour Necrosis Factor α, tumour necrosis factor, alpha, tumour necrosis factor, α	Extracellular Space	Cytokine
TP53	bbl, BCC7, bfy, bhy, LFS1, p44, P53, P53 cellular tumour antigen, p53 tumor suppressor, transformation related protein 53, TRP53, tumor protein p53, tumour protein p53	Nucleus	transcription regulator
TRIM23	6330516O20Rik, AI450195, ARD1, ARFD1, RNF46, tripartite motif-containing 23	Nucleus	enzyme
UBC	2700054O04Rik, AI194771, HMG20, Polyubiquitin, Rps27a, TI-225, Uba52, Ubb, UBIQUITIN C	Cytoplasm	enzyme
UBQLN4	A1U, A1Up, AI663987, C1orf6, CIP75, RGD1308273, UBIN, ubiquilin 4	Cytoplasm	other
USP12	LOC684318, UBH1, ubiquitin specific peptidase 12, USP12L1	Cytoplasm	peptidase
USP33	9830169D19RIK, AA409780, ubiquitin specific peptidase 33, VDU1	Cytoplasm	peptidase
USP45	3110003C05Rik, 4930550B20RIK, AI843191, GCAP7, ubiquitin specific peptidase 45, ubiquitin specific petidase 45	Other	peptidase
ZBTB7B	C2H2-Type Domain Containing Protein, CKROX, hcKROX, [45]THPOK, ZBTB15, ZFP-67, zinc finger and BTB domain containing 7B, ZNF857B	Nucleus	transcription regulator

III.4 Conclusions

Functions of the PCDHs family have not yet been described in detail. Our approach is based on the 'guilt by association' concept (introduced on Chapter II) implying a correlation between mutually interacting proteins and cooperation for the same function. Identifying interactor partners of different members of the δ-PCDH family might elucidate the involvement of these PCDHs in specific pathways.

First, we performed a screening to identify a list of candidate interactors. After a few attempts with Yeast 2-hybrid and classical MAPPIT approaches, the array MAPPIT screening method turned out to be the most appropriate test for our research questions. Results from these screenings were confirmed in binary (one-to-one) MAPPIT experiments. Due to a significant background signal in case of PCDHs baits, the resulting list was not as precise as hoped for. Anyway, we could identify in the top of our ranking list both known interaction partners of PCDHs as well as hits from previous experiments, confirming the validity of at least some array MAPPIT results. This is the case for PP1CA, a known interactor of PCDH11X scoring in position 15 of the PCDH11X array, and for PDLIM7, listed second in the same experiment, and previously identified as new candidate interaction partner in binary MAPPIT experiments and co localization test. The final short list includes what has been considered the most reliable and interesting candidates to reach the scope. An example is represented by three proteins, belonging to the MAX-MAD pathway, which are highly associated with cancer. The analysis of the re-test experiments confirmed 71 of the 288 candidates proposed by the array MAPPIT experiments (Identified by (+) in the column "retest", table 11). After additional cross test experiments, we created a short list of candidate interactors, which were positive in confirmation test using the original PCDH as bait but often also being positive in interaction with other members of the δPCDH family. 48 candidates were included in this final list (pink in table 11).

The list of the 48 candidates was investigated for diseases and common functions (Ingenuity Pathway Analysis). We identify 12 different association groups where at least 15 molecules of the 48 were involved. Details are shown in **Table 18**. Our results indicate that every group of 35 to 15 molecules proposed to be interaction partner of PCDHs family member, play a role in cancer.

Table 18 Clusters of candidates interaction partners of PCDHs are involved in cancer.

Diseases or Functions Annotation	Molecules	# Molecules
Cancer	ACOT8,CUL5,DPYSL5,DTX3L,EPB42,FHL3,GGN,GTPBP4,HEY1,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SH3D19,SKIL,SP6,STAU1,SYT14,TRIM23,TTLL1,UBTD2,UNC13D,USP12,USP33,USP46,ZBTB7B	36
tumorigenesis of tissue	ACOT8,CUL5,DPYSL5,DTX3L,EPB42,GGN,GTPBP4,HEY1,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SH3D19,SKIL,SP6,STAU1,SYT14,TRIM23,TTLL1,UBTD2,UNC13D,USP12,USP33,USP46,ZBTB7B	35
malignant solid tumor	ACOT8,CUL5,DPYSL5,DTX3L,EPB42,FHL3,GGN,GTPBP4,HEY1,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SKIL,SP6,STAU1,SYT14,TRIM23,TTLL1,UNC13D,USP12,USP33,USP46,ZBTB7B	34
epithelial cancer	ACOT8,CUL5,DPYSL5,DTX3L,EPB42,GGN,GTPBP4,HEY1,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SKIL,SP6,STAU1,SYT14,TRIM23,TTLL1,UNC13D,USP12,USP33,USP46,ZBTB7B	33
digestive system cancer	ACOT8,CUL5,DPYSL5,DTX3L,EPB42,GGN,GTPBP4,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SKIL,SP6,STAU1,SYT14,TRIM23,TTLL1,UNC13D,USP33,USP46,ZBTB7B	31
abdominal cancer	ACOT8,CUL5,DPYSL5,DTX3L,EPB42,GGN,GTPBP4,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SKIL,SP6,STAU1,SYT14,TRIM23,TTLL1,UNC13D,USP33,USP46,ZBTB7B	31
Adenocarcinoma	ACOT8,CUL5,DPYSL5,DTX3L,EPB42,GGN,GTPBP4,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SKIL,SP6,STAU1,SYT14,TRIM23,TTLL1,USP33,USP46,ZBTB7B	30
gastrointestinal tract cancer	CUL5,DPYSL5,DTX3L,EPB42,GGN,GTPBP4,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SP6,STAU1,TRIM23,TTLL1,USP33,ZBTB7B	26
abdominal adenocarcinoma	CUL5,DTX3L,EPB42,GGN,GTPBP4,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NKD2,PAX8,PDE1B,PDLIM7,RAB34,RFTN2,SKIL,STAU1,SYT14,TRIM23,TTLL1,USP33,ZBTB7B	25
malignant neoplasm of large intestine	CUL5,DPYSL5,DTX3L,EPB42,GGN,GTPBP4,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NFIA,NKD2,PAX8,PDE1B,PDLIM7,RFTN2,SP6,TRIM23,TTLL1,USP33,ZBTB7B	24
breast or colorectal cancer	CUL5,DTX3L,EPB42,FHL3,GTPBP4,HOXB7,KLF14,KLF4,MAX,MGA,MXD3,MXI1,NKD2,PAX8,PDE1B,PDLIM7,TRIM23,USP12,ZBTB7B	19
liver carcinoma	ACOT8,CUL5,DTX3L,EPB42,GGN,HOXB7,KLF14,MGA,MXD3,NKD2,PDE1B,SKIL,TRIM23,UNC13D,USP33	15

To further prioritize the members of this selection list, several Ingenuity Pathway analyses have been performed, and different lists (i.e. common or exclusive candidates) were compared to identify correlations and collaborations in specific pathways. Our results show that the candidate interactors in our priority list belong to different pathways and are involved in many biological

functions. At this stage we could not identify any strong indication concerning the function of PCDHs: our results only show that many mechanisms, involved in a variety of functions, might host members of the δPCDH family, unfortunately without offering any firm conclusion concerning the roles that PCDHs might play in cells.

Few candidates from the list were then selected for further molecular investigation. MAX, MXD3 (MAD3) and MXI1, as already mentioned, belong to the same pathway strongly related to cancer. Our selection included also PDLIM7, a candidate previously identified and still to be confirmed independently; a small group of ubiquitin-related proteins; FHL3, an interactor of FHL2 previously proposed to interact with PCDH11Y in a pull down assay; DMRTC1 and ZBTB7, common candidates for the three PCDHs; a short list of actin-related proteins suggested by the PCDH literature; PSMC1, proposed but not always confirmed in MAPPIT but interesting given literature indications. These candidates form together what we call the priority list: the following experiments (Chapter IV) of confirmation in biochemical assays were performed on the basis of this priority list.

III.5 References

1. van Roy, F. & Berx, G. The cell-cell adhesion molecule E-cadherin. *Cell. Mol. Life Sci.* **65,** 3756–88 (2008).

2. Hulpiau, P. & van Roy, F. Molecular evolution of the cadherin superfamily. *Int. J. Biochem. Cell Biol.* **41,** 349–69 (2009).

3. Redies, C., Vanhalst, K. & Roy, F. delta-Protocadherins: unique structures and functions. *Cell Mol Life Sci* **62,** 2840–2852 (2005).

4. Miyake, K. *et al.* The protocadherins, PCDHB1 and PCDH7, are regulated by MeCP2 in neuronal cells and brain tissues: implication for pathogenesis of Rett syndrome. *BMC Neurosci* **12,** 81 (2011).

5. Williams, E. O. *et al.* Delta Protocadherin 10 is Regulated by Activity in the Mouse Main Olfactory System. *Front. Neural Circuits* **5,** 9 (2011).

6. Caselli, R. *et al.* Retinoblastoma and mental retardation microdeletion syndrome: clinical characterization and molecular dissection using array CGH. *J. Hum. Genet.* **52,** 535–542 (2007).

7. Priddle, T. H. & Crow, T. J. Protocadherin 11X/Y a Human-Specific Gene Pair: an Immunohistochemical Survey of Fetal and Adult Brains. *Cereb. Cortex* **23,** 1933–1941 (2012).

8. Carrasquillo, M. M. *et al.* Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease. *Nat. Genet.* **41,** 192–8 (2009).

9. Speevak, M. D. & Farrell, S. A. Non-syndromic language delay in a child with disruption in the Protocadherin11X/Y gene pair. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **156B,** 484–9 (2011).

10. Kasnauskiene, J. *et al.* A single gene deletion on 4q28.3: PCDH18 – A new candidate gene for intellectual disability? *Eur. J. Med. Genet.* **55,** 274–277 (2012).

11. Dibbens, L. M. *et al.* X-linked protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment. *Nat. Genet.* **40,** 776–81 (2008).

12. Koppelman, G. H. *et al.* Identification of PCDH1 as a novel susceptibility gene for bronchial hyperresponsiveness. *Am. J. Respir. Crit. Care Med.* **180,** 929–35 (2009).

13. Dror, N. *et al.* Identification of IRF-8 and IRF-1 target genes in activated macrophages. *Mol. Immunol.* **44,** 338–46 (2007).

14. Li, X. *et al.* IKKalpha, IKKbeta, and NEMO/IKKgamma are each required for the NF-kappa B-mediated inflammatory response program. *J. Biol. Chem.* **277,** 45129–40 (2002).

15. Vazquez-Cintron, E. J. *et al.* Protocadherin-18 is a novel differentiation marker and an inhibitory signaling receptor for CD8+ effector memory T cells. *PLoS One* **7,** e36101 (2012).

16. Koneru, M., Schaer, D., Monu, N., Ayala, A. & Frey, A. B. Defective proximal TCR signaling inhibits CD8+ tumor-infiltrating lymphocyte lytic function. *J. Immunol.* **174,** 1830–40

(2005).

17. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–6 (2008).

18. Yu, J. S. *et al.* PCDH8, the human homolog of PAPC, is a candidate tumor suppressor of breast cancer. *Oncogene* **27,** 4657–65 (2008).

19. Imoto, I. *et al.* Frequent silencing of the candidate tumor suppressor PCDH20 by epigenetic mechanism in non-small-cell lung cancers. *Cancer Res.* **66,** 4617–26 (2006).

20. Narayan, G. *et al.* Protocadherin PCDH10, involved in tumor progression, is a frequent and early target of promoter hypermethylation in cervical cancer. *Genes. Chromosomes Cancer* **48,** 983–92 (2009).

21. Ying, J. *et al.* Frequent epigenetic silencing of protocadherin 10 by methylation in multiple haematologic malignancies. *Br. J. Haematol.* **136,** 829–32 (2007).

22. Yu, J. *et al.* Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* **136,** 640–51.e1 (2009).

23. van Roy, F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nat. Rev. Cancer* **14,** 121–34 (2014).

24. Terry, S. *et al.* Protocadherin-PC promotes androgen-independent prostate cancer cell growth. *Prostate* **66,** 1100–13 (2006).

25. Vanhalst, K., Kools, P., Staes, K., van Roy, F. & Redies, C. delta-Protocadherins: a gene family expressed differentially in the mouse brain. *Cell Mol Life Sci* **62,** 1247–1259 (2005).

26. Nakao, S., Platek, A., Hirano, S. & Takeichi, M. Contact-dependent promotion of cell migration by the OL-protocadherin-Nap1 interaction. *J. Cell Biol.* **182,** 395–410 (2008).

27. Eyckerman, S. *et al.* Design and application of a cytokine-receptor-based interaction trap. *Nat. Cell Biol.* **3,** 1114–9 (2001).

28. Uyttendaele, I. *et al.* Mammalian protein-protein interaction trap (MAPPIT) analysis of STAT5, CIS, and SOCS2 interactions with the growth hormone receptor. *Mol Endocrinol* **21,** 2821–2831 (2007).

29. Lievens, S. *et al.* Design of a fluorescence-activated cell sorting-based Mammalian protein-protein interaction trap. *Methods Mol. Biol.* **263,** 293–310 (2004).

30. Ulrichts, P., Peelman, F., Beyaert, R. & Tavernier, J. MAPPIT analysis of TLR adaptor complexes. *FEBS Lett* **581,** 629–636 (2007).

31. horfeome v5.1. at <http://horfdb.dfci.harvard.edu/hv5/>

32. Montoye, T., Lemmens, I., Catteeuw, D., Eyckerman, S. & Tavernier, J. A systematic scan of interactions with tyrosine motifs in the erythropoietin receptor using a mammalian 2-hybrid approach. *Blood* **105,** 4264–71 (2005).

33. Eyckerman, S., Broekaert, D., Verhee, A., Vandekerckhove, J. & Tavernier, J. Identification of the Y985 and Y1077 motifs as SOCS3 recruitment sites in the murine leptin receptor.

FEBS Lett. **486,** 33–7 (2000).

34. Simicek, M. *et al.* The deubiquitylase USP33 discriminates between RALB functions in autophagy and innate immune response. *Nat. Cell Biol.* **15,** 1220–30 (2013).

35. Moerkerke, B. & Goetghebeur, E. Selecting 'significant' differentially expressed genes from the combined perspective of the null and the alternative. *J. Comput. Biol.* **13,** 1513–31 (2006).

36. Caligiuri, M. *et al.* MASPIT: three-hybrid trap for quantitative proteome fingerprinting of small molecule-protein interactions in mammalian cells. *Chem. Biol.* **13,** 711–22 (2006).

37. Lievens, S. *et al.* Array MAPPIT: high-throughput interactome analysis in mammalian cells. *J Proteome Res* **8,** 877–886 (2009).

38. Lamesch, P. *et al.* hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* **89,** 307–15 (2007).

39. Lievens, S., Vanderroost, N., Defever, D., Van der Heyden, J. & Tavernier, J. ArrayMAPPIT: a screening platform for human protein interactome analysis. *Methods Mol. Biol.* **812,** 283–94 (2012).

40. Durick, K., Gill, G. N. & Taylor, S. S. Shc and Enigma are both required for mitogenic signaling by Ret/ptc2. *Mol. Cell. Biol.* **18,** 2298–308 (1998).

41. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322,** 104–10 (2008).

42. Xu, Y. *et al.* PCDH10 inhibits cell proliferation of multiple myeloma via the negative regulation of the Wnt/β-catenin/BCL-9 signaling pathway. *Oncol. Rep.* **34,** 747–54 (2015).

43. Ying, J. *et al.* Functional epigenetics identifies a protocadherin PCDH10 as a candidate tumor suppressor for nasopharyngeal, esophageal and multiple other carcinomas with frequent methylation. *Oncogene* **25,** 1070–80 (2006).

44. Li, Z. *et al.* Epigenetic inactivation of PCDH10 in human prostate cancer cell lines. *Cell Biol. Int.* **35,** 671–6 (2011).

45. Li, Z. *et al.* Role of PCDH10 and its hypermethylation in human gastric cancer. *Biochim. Biophys. Acta* **1823,** 298–305 (2012).

Chapter IV. BIOCHEMICAL CONFIRMATION AND PATHWAY ANALYSIS

IV.1 Biochemical pull down experiments

IV.1.1 Strategy and initial pilot experiments

To confirm the results obtained from the MAPPIT assay (Chapter III), we performed biochemical pulldown experiments. Immunoprecipitation (IP) is the technique which allows the precipitation of a target protein from a solution, generally a cell lysate, using an antibody that is target-specific. We performed experiments using Dynabeads to capture the antibody. This tool allows simple, mild, but efficient washing steps and the resulting immunoprecipitate can be analyzed by western blotting or mass spectrometry directly after elution from the washed beads. To identify interacting proteins, plasmids encoding EYFP-tagged full-length δPCDHs were transfected into Hek293T cells. Molecular complexes were then isolated by using a mouse anti-GFP antibody coupled to magnetic beads (Dynabeads M-280 Sheep anti-mouse IgG, Invitrogen), which acts as the solid support for binding Ig or target molecules. The aim was to identify by mass spectrometry (MS) the proteins pulled down together with EYFP tagged PCDHs.

Before starting large scale experiments, we tested the Dynabead-based pulldown procedure on the confirmed interaction between PCDH11X and DYNLT1 (Uta Brunner, not published). Next, we performed a pull down experiment on PCDH11X. The primary antibody used was a commercial anti-GFP (Invitrogen). Western blot data were promising as the GFP-fused PCDH11X was clearly detectable (**Figure 60**) although CBB (Coomassie Brilliant Blue) staining was uninformative.

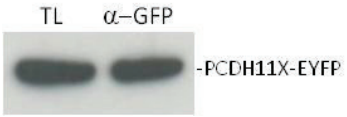


Figure 60 Pilot pull down experiment.
Western blot with anti-GFP antibody for pull down experiments for PCDH11X-EYFP. TL=total lysate

Full-length PCDH11X, PCDH11Y, PCDH10, PCDH18 and PCDH9, all EYFP tagged, were overexpressed by transient transfection in Hek293Tcells and localized by fluorescence microscopy. The membrane localization was not clear as expected, above all for PCDH10. Localization experiments of transiently overexpressed PCDH10 were performed also in other cell lines as shown in **Figure 61.** We noticed protein aggregates in the cytoplasm, indicating an unexpected localization of PCDHs. Our data indicate a possible localization of a subpopulation of PCDH10 at the perinuclear regions in HeLa cells, suggesting an internalization subsequent an interaction with a dynein motor complex, but this event upon ectopic overexpression can also be due to the fact that the transfected cells are not able to

manage such a high production of artificially induced protein. We also attempted to establish stable cell lines with more physiological expression levels of PCDHs in order to minimize possible overexpression artifacts. This stable cell line showed to be suitable for other experiments but didn't perform well for CoIP which was then performed in Hek293Tcells transiently transfected for the different PCDHs and relevant candidates.

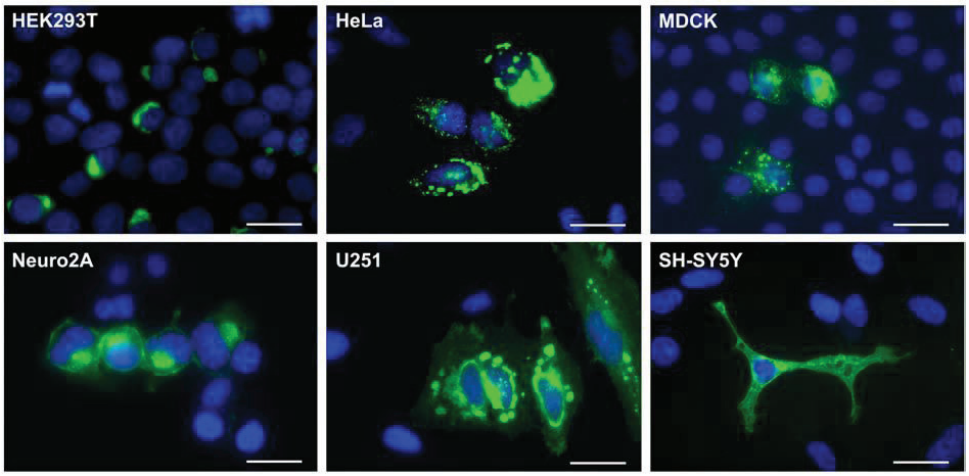


Figure 61 PCDH10 accumulates in the cytoplasm if transiently transfected in various cell lines
EYFP-PCDH10 was transiently overexpressed in a number of cell lines. The expected membrane localization was not clearly observed in every cell line where protein accumulations were noticed. (Picture by I. Kahr, not published).

IV.1.2 Confirmation CoIP

CoImmunoprecipitation tests have been performed for different PCDHs and a short list of candidate interactors. As discussed in Chapter III, the long MAPPIT-based list of candidates was reduced to a rather short priority list. We performed experiments in Hek293T cells transiently transfected with EYFP tagged PCDHs.

IV.1.2.1 Materials and methods

CoImmunoprecipitation: CoIP experiments were performed in Hek293Tcells, cultured in Dulbecco's modified Eagle's medium (DMEM) +10% fetal bovine serum. The day before transfection, 300.000 cells were seeded in each standard 60 mm petri dish. To confirm interactions, GFP-tagged δPCDHs were co-transfected with Flag-tagged candidates in Hek293T cells. Cells were transfected using CaP precipitation method as previously described. Medium was changed after 4/6 hours. Cells were lysed 24h after transfection in IP lysis buffer (150 mM NaCl, 2 mM EDTA, 25 mM Tris pH 7.4, 0.5 % Triton-X100, complete protease inhibitor mix); lysate was scraped off and transferred in a cold tube, incubated while rotating for 10 minutes and centrifuged for 20 minutes at 4°C at high speed. Protein concentration was determined with the DC protein assay (Bio-Rad) and 40 µg per sample were kept as input. CoIP was performed using Dynabeads M-280 Sheep anti-mouse IgG (Invitrogen) according to the manufacturer's instructions. Dynabeads were re-suspended by vortexing and 50 µl Dynabeads suspension per IP were transferred to a tube, placed on the magnet and the supernatant was removed. 1 µg antibody diluted in 200 µl PBS/ 0.02% Tween20 (PBS/T) was added and incubated with rotation for 10 min at RT or 2h at 4°C. The complexes were washed once and suspended in 200 µl PBS. The Dynabeads were washed twice with IP lysis buffer and incubated with 1 mg cell lysate for 10 minutes (up to 2 hours) at RT with rotation. The Dynabeads-antibody-antigen complex was washed three times using 1 ml IP lysis buffer. After the second wash, the beads were transferred to a new tube. After the final wash, the tube was placed on the magnet and the buffer was removed from the tube (without letting the beads dry out completely). 50 µl of 1XSDS loading buffer (50 mM Tris pH 6.8, 2 % SDS, 0.1 % bromophenol blue, 10% glycerol, 100 mM β-mercaptoethanol) were added and gently pipetted up and down to re-suspend the Dynabeads-antibody-antigen complex. After incubation for 10 min at 95°C, the tubes were placed onto the magnet and the eluate was loaded on an SDS-PAGE gel for Western blot analysis.

SDS-PAGE and Western blotting: 40 µg of total protein cell lysate and 50 µl of eluate from the CoIP were loaded on a one-dimensional SDS-PAGE gel. Proteins were transferred to PVDF membranes

(Millipore) which was blocked with 5% non-fat dry milk in TBS -0.1 % Tween-20. Membranes were incubated with primary antibodies overnight at 4°C. After several washing steps in TBS-Tween and TBS, membranes were incubated for 1 hour with secondary horseradish peroxidase (HRP)-conjugated antibodies. Detection was performed using the ECL detection system (Amersham GE healthcare). Primary Antibodies: for CoIP, Anti-GFP 3E6 Mouse Invitrogen, Cat.-No. A11120; for WB, Anti-GFP Mouse Roche, Cat.-No. 11814460001 1/1000 Anti-FLAG Mouse Sigma, Cat.-No. F3165 1/1000. Secondary Antibodies: for WB, Anti-mouse HRP Sheep GE Healthcare, Cat.-No. NA931V 1/3000.

IV.1.2.2 Results

We initially performed experiments for PCDH10 and PCDH11X and co-transfected EYFP-tagged PCDH constructs with flag-tagged interaction candidates encoding plasmids. To start we optimized the protocol by testing a known interaction partner, DYNLT1 (**Figure 62**).

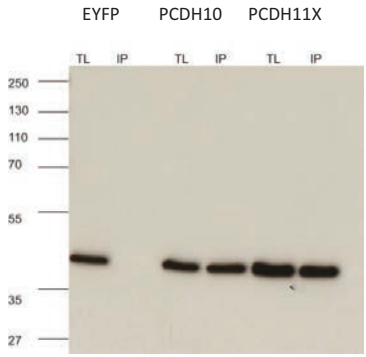


Figure 62 DYNLT1 coimmunoprecipitates with PCDH10 and PCDH11X
DYNL1 flag-tagged and PCDH10-EYFP or PCDH11x-EYFP or EYFP alone (as control) were expressed in Hek293Tcells. IP was performed with αGFP antibody and WB with αflag antibody.
DYNLT1 co-precipitates with PCDH10 and with PCDH11X as expected but not with EYPF.

We started the confirmation experiments with the MAX-MAD group of candidates as they have an established role in cancer. We could clearly show a strong interaction of the MAX protein with both PCDHs (**Figure 63**). Our blots confirmed also the interaction of PCDH10 with MAD3 (MXD3), and more modestly with MXI1. Surprisingly, we could not show the interaction with MAD3 and MXI1 for PCDH11, even though members of this MAX-MAD group turned out to be positive in all kinds of our PCDH-MAPPIT experiments, and were not identified in MAPPIT experiments with other baits (personal communication by Sam Lievens). Furthermore, we performed CoIP tests for PCDH1 and PCDH7 to investigate whether MAX could be a common interaction partner and this indeed turned out to be the case. The experiment was performed for the long (iso1) and the short (iso2) isoforms

of MAX which were both indicated as putative interaction partners in MAPPIT experiment. A construct missing the gp130 (present in the MAPPIT plasmids) was used to exclude an aspecific interaction for iso1. The results of CoIP experiments together with our MAPPIT findings identified MAX as a candidate interaction partner for PCDH10, PCDH11X, PCDH9, PCDH1 and PCDH7. Interaction with MAD3 could be confirmed only for PCDH10 despite the positive MAPPIT results for PCDH11X and PCDH9 as well as PCDH10 (**Figure 63**). Both irrelevant antibody and EYFP antibody resulted in a precipitation of MAD3 and MXI1 for PCDH11X. Thus, given the aspecificity of the signals, these results are inconclusive and a different analysis system might be used to confirm the interactions suggested by our MAPPIT experiments.

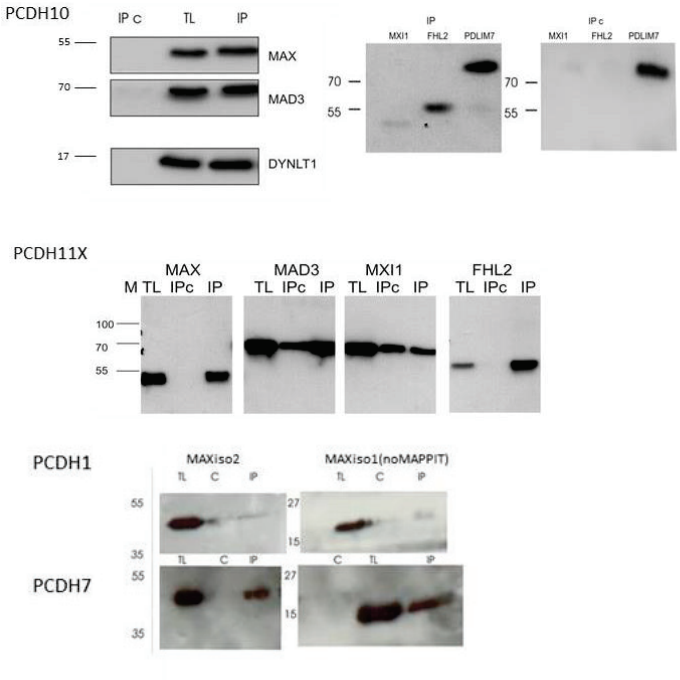


Figure 63 Coimmunoprecipitation of different PCDHs with putative interaction partners
WB analysis of lysate from Hek293T cells transiently transfected with PCDH10-EYFP, PCDH11-EYFP, PCDH1-EYFP or PCDH7-EYFP and MAX-flag, MAD3-flag, MXI1-flag, FHL2-flag and PDLIM7-flag. CoIP performed with antibody anti GFP to precipitate PCDHs EYFP tagged (IP) or with irrelevant antibody as control (IPc). TL=total lysate. Known interaction between PCDH10 and DYNLT1 has been used as positive control. WB performed with anti-flag antibody.

Both PCDH10 and PCDH11X interact with FHL2 (**Figure 63**). FHL2 has been shown to interact with PCDH11Y and it has been found positive in our binary MAPPIT (details in Chapter III). PDLIM7, another of our high ranking interaction candidates, immunoprecipitates with PCDH11X but the

negative control (IP with irrelevant antibody) shows an equally strong signal which implies that the interaction seen in this CoIP can be aspecific. Beside those first priority candidates we checked the interaction of PCDHs also with a few more candidate interactors from the array.

CUL5 and PSMC1 became interesting interaction candidates on the basis of literature. Indeed, a connection between PCDH10 and PSD-95 (post synaptic density protein 95) has been recently reported[1]. Tsai *et al.* showed that Pcdh10 plays a role in MEF2-induced synapse elimination and functions by delivering ubiquitinated PSD-95, a critical synaptic scaffolding molecule, to the proteasome. CUL5 is one of the proteins in the list of our interaction candidates and the Cullin protein family plays a role in the post-translational ubiquitin modification of cellular proteins. Our list of candidates included PSMC1, a proteome subunit, as well: we wanted to confirm these interactions to demonstrate any role of this protein in the binding of PCDH10 to the proteasome. CoIP experiments confirmed the interaction of CUL5 with PCDH10 but not with PCDH11X, while PSMC1 interacted with neither PCDH10 nor PCDH11X in CoIP experiments (**Figure 64**).

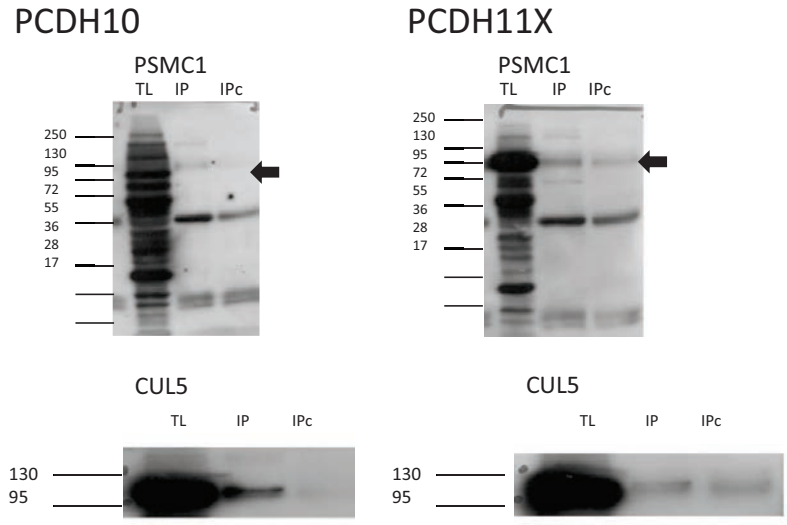


Figure 64 Coimmunoprecipitation of PCDH10 or PCDH11X with PSMC1 and CUL5
WB analysis of lysate from Hek293T cells transiently transfected with PCDH10-EYFP or PCDH11-EYFP and CUL5 or PSMC1 flag tagged. Lysate CoIP performed with anti-GFP (IP) antibody or with an irrelevant antibody as control (IPc). TL=total lysate. WB performed with anti-flag antibody.

CoIP experiments have been performed also for different members of the USP family with the same protocol. As it is shown in **Figure 65** and in **Figure 66**, USP21 co-immunoprecipitates with the 3 investigated PCDHs but only with PCDH9 in a specific way. PCDH9 interacts with USP33, and both

PCDH10 and PCDH11X co-immunoprecipitate with USP46. The interaction with CUL5, specific for PCDH10, was used as control.

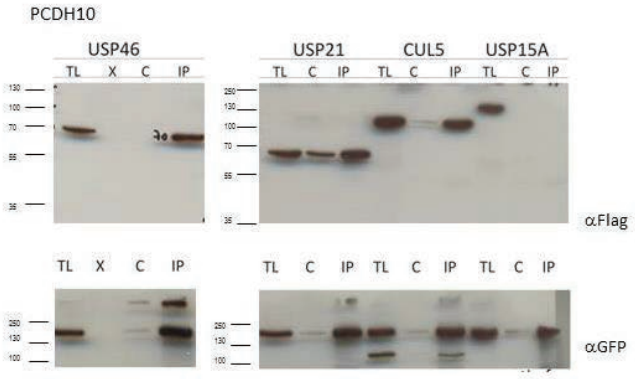


Figure 65 Coimmunoprecipitation experiments of PCDH10 with USP family members
WB analysis of CoIP experiments performed in lysates of Hek293T cells transiently co-transfected with EYFP tagged PCDH10. IP= αGFP antibody, WB= αFlag or αGFP antibody as indicated; TL, total lysate; X, empty line; C, negative-control antibody; IP, specific antibody (= αGFP).

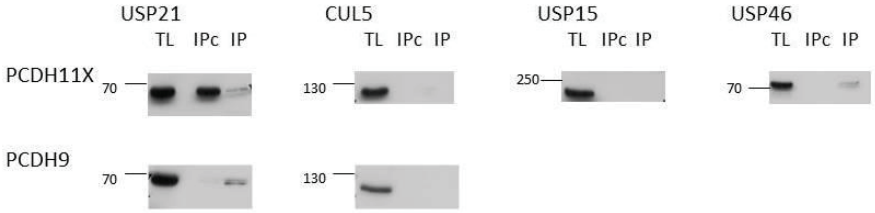


Figure 66 Coimmunoprecipitation experiments of PCDH11X and PCDH9 with USP family members
WB analysis of CoIP experiments performed in lysates of Hek293T cells transiently co-transfected with EYFP tagged PCDH9, PCDH10 or PCDH11X together with flag tagged putative interaction partner belonging to USP family members as indicated on top. CoIP performed with anti-GFP antibody to precipitate PCDHs EYFP tagged (IP) or with irrelevant antibody as control (IPc). TL=total lysate. WB performed with anti-flag HRP-coupled antibody.

IV.2 Interaction candidates descriptions

IV.2.1 MAX-MAD family

MYC-associated protein X (MAX) is a ubiquitous and constitutively expressed protein, which plays a central role in the control of MYC. MYC is a very well-studied nuclear phosphoprotein involved in cell cycle progression, apoptosis and cellular transformation[2]. Together with MYCN and MYCL proteins, MYC is altered in function and expression in most human cancers[3]. For this reason many studies have been carried out to investigate the functional relevance of MYC and to identify MYC interacting proteins. These studies first identified the presence at the extreme C terminus of a motif previously identified in many transcription factors, the basic-helix-loop-helix-leucine zipper (bHLHZip). This motif has been shown to be responsible for protein dimerization and DNA binding. To investigate if MYC behaves as the other transcription factors, PPIs studies were carried out, and, indeed, the prediction were confirmed: in this way MAX was identified as a MYC interaction partner[2–5]. MAX, as MYC, belongs to the bHLHZip protein family of transcriptor factor and is expressed in two isoforms which migrate as 21 and 22 kDa[5]. The short one is a 151 amino acid-long polypeptide while the long one is a 160 amino acid chain. The differential splicing occurs thanks to a sequence encoding nine amino acids in the N-terminus of MAX. Both isoforms are highly stable and independent from the cell cycle stage[6]. The short and the long isoforms take parts in the DNA binding, cell cycle progression, and apoptosis process in different ways. The strongest difference is the DNA binding capacity, which obviously has an impact on the protein's effect on the target genes[6]. MAX was shown to interact with all MYC family proteins: the heterocomplex is able to recognize the DNA sequence CACGTG (an E-boxe sequence). The MYC network is very complex and requires many control mechanisms to coordinate all the cellular activities where it is involved.

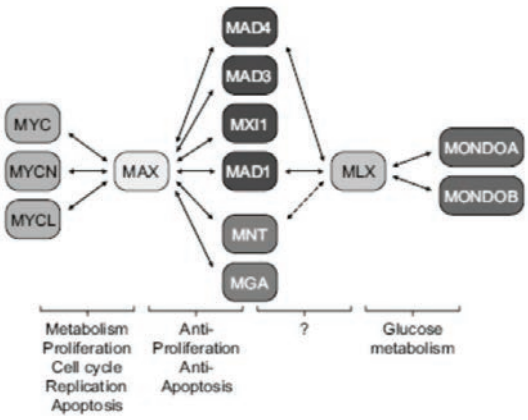


Figure 67 The Myc/Max/Mad network.
The figure shows the components of the Myc/Max/Mad network. Known interactions (by the HLHZip dimerization domains) are shown with arrows. The network functions as regulator of many events of the cell cycle and metabolism and not all mechanisms have been elucidated yet. Figure from[3].

An additional regulation is given by the interaction with MAX at the level of the formation of the heterodimers[4]. This regulation occurs due to a family of proteins which competes with MYC for the binding to MAX. The first component of the family to be described were MAD1[7] and MXI1[8]. Both these proteins bind to MAX with the same affinity of MYC. Few years later, Hurlin and colleagues[9] identified other two member of the MAD family: MAD3 and MAD4. The four proteins behave as MYC: they only show a weak homodimerization and DNA binding ability. The MAD-MAX heterodimer act as a transcription repressor[2], despite using the same binding site of MYC-MAX complex. Furthermore MNT and MGA were described as MAX-interacting protein, but they are poorly related to the MAD family, even if they show some MAD-like functions[3] (**Figure 67**).

The MAPPIT assay carried out for PCDHs indicated some members of the MYC/MAX/MAD pathway as candidate interaction partners for different PCDHs. This was the only case in which different proteins of the same pathway turned out to be positive in our screening. Furthermore, as already mentioned, the deregulation of MYC/MAX/MAD pathway plays such an important role in the genesis of human cancer that it becomes the main focus of our further study. Interestingly, despite the consistent number of MAPPIT arrays performed in the lab of Prof. Tavernier, those candidates appear to be unique and specific for PCDHs. MAX turned out to be a candidate interaction partner for all the 3 studied PCDHs. The role of MAX in cancer has been shown in different studies. For example, MAX has been defined a tumor suppressor gene which causes hereditary pheochromocytoma (PPC)[10]: mutation in the gene were detected in patients and localized in most of the cases in the conserved amino acid of the bHLHZip region. If the ability of binding is lost, MAX cannot antagonize anymore the MYC dependent cell transformation[11]. This study, together with others, shows the importance of the MYC/MAX/MAD network in the development of neuronal crest tumor. The same has been shown for NMYC and his role in the development of neuroblastoma[12]. The members of the MAD family have been shown to suppress the development of cancer[13], to promote differentiation *in vivo* and block cellular growth, while MNT can be considered a regulator of MYC and MGA is suggested to silence MYC responsive genes In quiescent cells[14]. It appears that despite some differences in action, the members and relatives of the MAD family, behave in the same way giving a putative functional redundancy even if it seems that they perform their functions in different timing due to differential expression induction[11]. MXI1 and MAD3 are generally expressed in the early phases of differentiation or in proliferating cells while MAD1 and MAD4 are expressed in the terminal stages of differentiation. Interestingly enough, the candidates belonging to the family which results positive in the MAPPIT experiments are involved in brain cancer. For examples *MXI1* which is localized in chromosome 10q24-q25 (a cancer hotspot), shows a deletion

often found in brain cancer (i.e. 80% of glioblastoma multiforme and 15% of primary medulloblastoma). Furthermore the reintroduction of MXI1 in a MXI1 missing human glioblastoma cell line results in growth reduction and clonogenicity, and shows a high percentage of cells with DNA degradation with accumulation of cells in the G2-M phase. This light on the role in the pathogenesis of human glioblastoma, suggests that MXI1 can be considered a tumor suppressor gene, as expected[15]. Moreover, mutation in MXI1 was observed also in prostate cancer [16].

MAD3 is a controversial and interesting candidate that MAPPIT proposed for PCDHs. If on the one hand it is similar to the other components of the family concerning structure and interaction with MAX, on the other hand it shows unique characteristics such as the expression profile during the different phases of the cell cycle. If MAX and MNT show a rather uniform expression and MAD1, MXI1, MAD4 are unregulated in quiescence and differentiation, MAD3 expression is associated with the last S-phase prior to terminal differentiation. This observation is surprising, given the role that MAD3 plays to inhibit proliferation and transformation.

The most intriguing feature of MAD3, making it unique in the family, comes from a study challenging the switch model of MYC/MAX to MAD/MAX. Yun and colleagues[17] demonstrated a new role of MAD3 in sonic hedgehog (shh)-dependent cerebellar granule neuron precursor (GNPs) proliferation and NMYC expression. The sonic hedgehog made by Purkinje neurons, regulates the vertebrate organogenesis and specifically controls the proliferations of GNPs (also) via NMYC during postnatal development. If Shh is over-activated, medulloblastoma can develop from the GNPs. They have shown that the inappropriate activation of MAD3 plays a role in tumor progression and stimulates proliferation. MAD3 is overexpressed in a mouse model of medulloblastoma and in many types of human brain tumors, including glioblastoma multiforme. In tumor, MAD3 is expressed during the complete cell cycle losing the S-phase exclusivity[17,18].

Considering our focus in cancer and the fact that PCDHs are mostly expressed in brain, those candidates look promising to be further investigated. The last member of the pathway which has been proposed as a candidate is MAD4.

Ingenuity pathway analysis (IPA) was used to further process the list of candidates obtained from the screening and in particular for MYC/MAX/MAD pathway, in order to speculate on the role of the interaction between the MYC/MAX/MAD pathway and the δPCDH family members. The analysis results in an even stronger confirmation of the role that the pathway plays in cancer, with a strong involvement in the nervous system functions.

IV.2.2 FHL2-FHL3

Four and a half LIM domains (FHL) proteins belong to the LIM-only protein family, a group of proteins characterized by the presence of a LIM domain, a double zinc finger motif with a consensus amino acid sequence CX2CX16–23HX2 (C/H) X2CX2CX16–23CX2(C/H/D) which mediates PPIs[19]. FHL2 proteins show an amino-terminal half LIM domain followed by a tandem of four full LIM domains[20]. FHL1, FHL2 and FHL3, three members of the family, show a homology of 50% in sequence identity and are mainly, but not exclusively, expressed in muscles. The family members regulate many transcription factors including SMAD proteins, β-catenin, FOXO1, SRF, AP-1, NFAT, MyoD, and the androgen receptor[21]. FHL2 is the better described member of the family and, as the other members, plays a role in the regulation of a wide range of cell functions, including cell cycle, signal transduction, gene expression, apoptosis and cell proliferation[21,22]. It has been indicated as a critical inducer of the epithelial-to-mesenchymal transition (EMT) and invasion.

Over 50 proteins have been shown to interact with FHL2. FHL2 homodimerizes and forms heterodimers with FHL3. Fimia and colleagues[23] have shown the interaction between FHL2 and FHL3 by Y2H technology, and independently Li *et al*.[24] obtained the same result via Y2H and FRET. FHL3 functions have not been well described yet, but an important role in the development and progression of breast cancer has been indicated[25]. Concerning the expression of FHL3, strong cytoplasmic staining has been shown in placental trophoblastic cells and smooth muscle cells. In male reproductive system, FHL3 is expressed in testis (cells of seminiferous ducts), prostate and epididymis glandular cells. Neuronal cells of hippocampus and lateral ventricles show moderate positivity in the nucleus. The clinical relevance of the FHL proteins is mostly due to their role in cancer. In a staining for 20 different types of cancer, FHL3 is positive in 27% of the cases and FHL2 in 33% (data from the Human Protein Atlas http://www.proteinatlas.org/). FHL2 is overexpressed in hepatocellular carcinoma if compared with normal liver, in epithelial ovarian and breast cancer[26,27], and in gastric and colon cancer[28]. Downregulation of FHL2 expression has been shown in rhabdomyosarcoma[29] and in primary prostate cancer[30], but another study reports that FHL2 expression is increased in prostate adenocarcinoma cells[31]. TGFβ1 stimulates Krüppel-like factor (KLF) 8, and KLF8-induced FHL2 activation has been recently identified as a critical mechanism underlying colorectal (and breast) cancer invasion and metastasis[32].

Expression of FHL2 in samples from patients operated for colorectal adenocarcinoma correlates with poor survival and nuclear expression of FHL2 in cancer cells associates with lymphatic metastasis in sporadic but not in HNPCC-associated colon cancer[33–35]. These results can be explained considering

the important role that FHL2 plays in EMT: FHL2 inhibits E-cadherin expression through interaction with Snail1 and promotes EMT in colon cancer[36].

The strong evidence for frequent epigenetic inactivation of PCDH10 in various human cancers, including gastric and prostate cancer, but not in matched normal tissues[37–40], made FHL2 an interesting candidate for our project. In a patent application for a method to treat or prevent hormone resistant prostate cancer using a siRNA specific (or another inhibitor) of PCDH-PC (PCDH11y) expression or activity, Buttyan and colleagues demonstrated an interaction between FHL2 and PCDH11Y in a pull down assay and showed that expression of these two proteins correlates with prostate cancer prognosis and outcome. The mechanism linking FHL proteins and neoplasia is very intriguing but not clear yet. Different expression levels of family members in normal and tumor tissues have been observed[41]. **Table 19** shows examples of FHL2 over-expressed or down-regulated in different types of cancer.

Table 19 FHL2 expression in cancer

Tissue	Normal tissue	Cancer
Breast	-	+
Colon	-	+
Lung	-	+
Muscle	+	-
Ovary	-	+
Prostate	+	-
Skin	?	+

FHL2 is a candidate PCDHs interaction partner, as shown from direct binary MAPPIT experiments for PCDH11X. Furthermore, FHL2 ranks in position 157 in the Array MAPPIT results for PCDH11X (25 as induction factor). FHL3 has been proposed as a candidate from the Array MAPPIT and confirmed in the retest, again for PCDH11X.

IV.2.3 PDLIM7

The structure of PDZ-LIM family members (or Enigma proteins) is characterized by the presence of an N-terminal PDZ domain and a series of C-terminal LIM domains. Enigma (PDLIM7), ENH, ZASP/Cypher, RIL, ALP, and CLP36 are members of the family and are cytoplasmic proteins. PDZ domain gets the name from the proteins where the domain was identified first (postsynaptic

density-95, discs large, and zo1 tight junction protein)[42]. It is a small, modular domain, which binds the consensus sequence (Ser/Thr)-X-(Val/Leu/Ile) of the target at the carboxyl terminus but also, for example, internal consensus sites and LIM domains[43–45]. Proteins which include the domain are often found in complexes at the cell-cell junctions. In particular, PDLIM7 is a 470 amino acid protein containing one PDZ and three LIM domains[46]; biologically it is considered an adapter: the PDZ domain can localize the LIM binding proteins to actin filaments. Most of the normal tissues are negative for PDLIM7 except a medium level expression in glial cells in cerebellar cortex, hippocampus and lateral ventricle. A medium level expression has been shown also in the heart myocytes. In a subcellular level, PDLIM7 is expressed at the actin filaments, but also in the nucleus and at the focal adhesion sites. The Human Proteomic Atlas includes staining for PDLM7 in 20 different kinds of cancer and positive signaling has been observed in 17% of the cases. The strongest signal is shown in glioma cells, but a clear signal was noticed also in some breast and endometrial cancers.

The interaction between PDLIM7 and δPCDHs has not been published yet but this candidate appears to be recurrent in our assays. PDLIM7 was first confirmed in a binary MAPPIT assay and in co-localization experiments for PCDH11X (**Figure 68**). The Array MAPPIT confirmed the interaction with PCDH11X: PDLIM7 is the second stronger candidate of the array MAPPIT list with a fold induction value of 119.

PDLIM7 is known to negatively regulate p53 through MDM2 and has been shown to promote tumor cell survival in mice[47]. In the light of this report, PDLIM7 was considered an intriguing candidate for our cancer related list. Furthermore, PDLIM7 interacts with the Receptor Tyrosine Kinase RET to

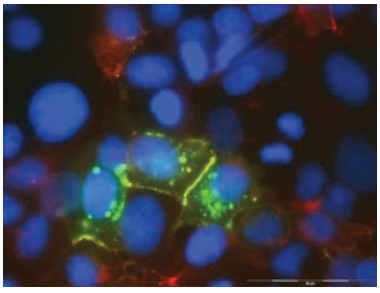


Figure 68 Co-localization of PCDH11X and PDLIM7
Hek293T cells transiently transfected with a plasmid encoding PCHD11X-EYFP and Flag-gp130-PDLIM7. Uta Brummer, not published.

trigger its function in mitogenic signalling[48]. It has been published that RET binds to clustered Pcdhs in differentiated neuroblastoma cells and it is required for stabilization and differentiation-induced phosphorylation of Pcdhs[49]. Actually, it appears that Ret and Pcdhs stabilize each other: while Ret regulates Pcdh-α and Pcdh-ϒ tyrosine phosphorylation and protein level, Pcdhs stabilize Ret in mouse central nervous system catecholaminergic cell line (CAD cells) and sympathetic neurons (**Figure 69**). This interaction is localized to the growth cones and synapses of neurons with demonstrated implications in axonal path finding and neuronal survival[50,51]. In the light of this, we believed that the putative interaction between PDLIM7 and δPCDHs was of great interest and we included the candidate in the short priority list.

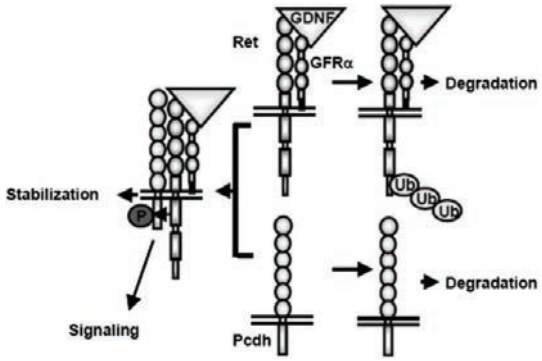


Figure 69 Model of Pcdh and Ret interaction and stabilization in CAD and sympathetic neurons[49].
Activation of Ret with GDNF/GFRα1 leads to ubiquitination and rapid degradation. Pcdhs not bound to Ret also undergo degradation. Activated Ret bound to Pcdh is stabilized. Pcdhs bound to activated Ret are stabilized and phosphorylated and might initiate downstream signaling.

IV.2.4 USPs family

Almost any protein undergoes posttranslational modifications, as the key mechanisms to increase proteomic diversity. There are many possible modifications. Two of the most frequent and relevant posttranslational modifications are phosphorylation and ubiquitination. Phosphorylation occurs on residues of serine, threonine or tyrosine and it is a reversible modification via phosphatases cleavage. Ubiquitination occurs on lysine residues, which are modified by the protein ubiquitin, a 76 amino-acid polypeptide that is highly conserved and abundant in eukaryotic cells[52,53]. It is a tag process: ubiquitinated proteins are recognized by the 26S proteasome for degradation.

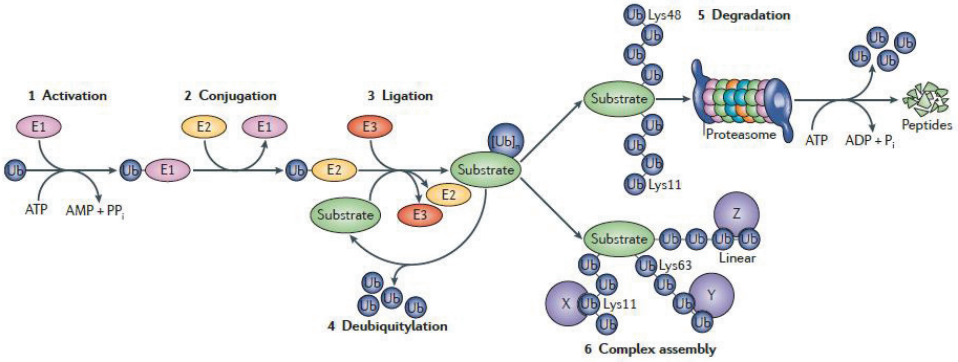


Figure 70 The enzymes and reactions of the UPS.
The ubiquitination and degradation of substrate proteins is achieved by a series of reactions mediated by the enzymes of the ubiquitin–proteasome system (UPS). In the activation reaction, ubiquitin is ATP-dependently transferred to an E1 (step 1). The activated ubiquitin is then transferred to an E2 enzyme in the conjugating reaction (step 2). Then the E2 enzyme carries the ubiquitin to the E3 enzyme, which is also known as an ubiquitin ligase (step 3). E3 is important not only because it covalently ligates ubiquitin to Lys residues on the substrate protein, but also because it mediates substrate specificity. This process of ubiquitin ligation may be repeated with a Lys of the ubiquitin protein itself serving as the substrate, which leads to the formation of a polyubiquitin chain on the target protein. Deubiquitinating enzymes may reverse substrate protein ubiquitination (step 4). Ligation of polyubiquitin has diverse biological consequences for the recipient protein. For example, Lys11- and Lys48-linked polyubiquitin chains serve as tags to target substrate proteins for proteasomal degradation (step 5). Conversely, linear, Lys63- and Lys11-linked chains promote the assembly of signaling complexes (step 6). X, Y and Z indicate ubiquitin-binding proteins. P_i, inorganic phosphate; PP_i, inorganic diphosphate; Ub, ubiquitin. Image is modified from[54].

During the ubiquitination process (**Figure 70**), ubiquitin molecules are attached to the acceptor protein in a complex three-step reaction, in the so called ubiquitin-proteasome pathway (UPP). First, an E1 ubiquitin activated enzyme starts the process activating the ubiquitin and transferring it in the active site, the amino acid cysteine, in an ATP-dependent reaction with the production of an intermediate substance known as ubiquitin-adenylate. A second enzyme, E2 (ubiquitin-conjugating enzyme), takes part in the process in the conjugation step. But the key enzyme in the process is the E3 (Ub-protein ligase): it acts as an adaptor, which binds activated E2 and protein substrate to allow isopeptide bond formation[d] [52,54,55].

[d] Isopeptide linkage: An amide bond that forms between a side-chain carboxyl group and amino group and is not present on the main chain of a protein. In the case of ubiquitination, isopeptide linkages form between the ε-nitrogen of Lys side chains and the C-terminus of the incoming ubiquitin, and constitute the basis of polyubiquitin chains[54].

There are three kinds of E3 Ub ligases distinguished on the basis of their E2-recruiting domains: RING,[e] U-box and HECT domains.

The process connects the last amino acid, glycine 76, of the ubiquitin molecule to a lysine on the substrate protein and can be repeated several times to build in this way chains of Ub. In humans there are two E1 enzymes, 37 Ub conjugating enzymes and more than 600 E3 Ub ligases[52]. When proteins undergo the ubiquitination modification they can have different biological destinies as it is shown in **Figure 70**, they can form complexes (20 types of Ub interaction domains have been identified and ubiquitin polymers can act as interaction site for proteins) or trigger proteasomal degradation. The UPP system is active for degradation in all tissues and for the majority of intracellular proteins. Another process is responsible for the degradation of extracellular and cell surface proteins and occurs in the lysosome[55].

There are about 100 human deubiquitinating enzymes (DUB) subdivided in 5 categories: ubiquitin C-terminal hydrolases, ubiquitin-specific processing proteases (USPs), Machado-Joseph disease protein domain proteases, ovarian tumor proteases (OTUs) and JAMM (JAB1/MPN/Mov34 metalloenzyme) motif proteases[56].

The biggest and better characterized subgroup of DUBs with over 60 members is the one of the cysteine protease USPs (ubiquitin-specific proteases). Their catalytic domain shows the so called histidine and cysteine boxes and their specific action is to hydrolyze bonds involving the carboxyl group at the C-terminal glycine residue of ubiquitin[53]. The reversibility of the ubiquitination process is an important factor to conserve the balance. That may explain why mutations that affect the function of USPs have been shown to occur in many cancers as well as in other disorders involving inflammation and the regulation of cell death[57].

The Array MAPPIT experiments for δPCDHs proposed some interaction candidates for different PCDHs which belong to the USPs family and these were confirmed in the binary assay and CoIP experiments. Between the candidates we could also identify CUL5 and PMSC1.

[e] RING domain: it includes a catalytic zinc finger like domain (Cys3HisCys4) able to chelate two zinc ions. U-box domain: a derived version of the RING-finger domain that lacks the hallmark metal-chelating residues of the latter 5 and 6 but is likely to function similarly to the RING-finger in mediating ubiquitin-conjugation of protein substrates[78]. HECT domain: it contains a catalytic Cys residue able to accept the charged Ub from E2[54].

IV.2.5 CUL5 & PMSC1

The Cullin family is a group of seven (CUL1, 2, 3, 4A, 4B, 5 and 7) hydrophobic proteins, which are used as scaffold proteins for the E3 ubiquitin ligase enzyme. The combination with RING proteins produces the so called Cullin-RING ubiquitin ligase (CRL). Cullins are then needed for ubiquitin-dependent protein degradation as already described in this dissertation.

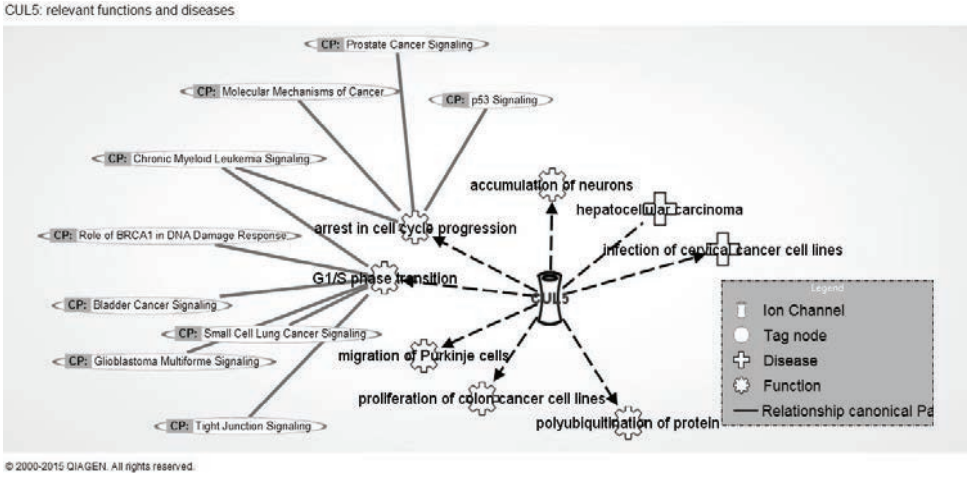


Figure 71 CUL5: relevant functions and diseases (Ingenuity® Systems, www.ingenuity.com)

In particular, Cullin5 (CUL5) is part of the SCF-like ECS (Elongin-Cullin 2/5-SOCS-box protein) E3 ubiquitin-protein ligase complexes. CUL5 has different functions beside protein degradation, including receptor activity and calcium channel activity[58] and it is involved in many biological activities including proliferation and migration of neuronal and cancer cells (**Figure 71**). *Cul5* is defined as a cancer related gene. The expression in many different tissues (including the male and female reproductive systems and the brain) and, above all, the important role played in many types of cancer, make CUL5 an interesting candidate partner for PCDHs. In our MAPPIT studies, CUL5 ranks in position 9 in the ARRAY list for PCDH11X and retests clearly confirmed this interaction. Cross test experiments showed a positive signal also for PCDH9. An interaction of CUL5 with PCDH10 was also seen but not always confirmed in MAPPIT, although PCDH10 can clearly pull down CUL5 in CoIP experiments.

PMSC1 (Proteasome 26S subunit, ATPase, 1) is another degradation-related candidate interactor, which was proposed by the Array MAPPIT experiments. Two complexes are cooperating to build the proteasome system: 4 rings of 28 non-identical subunits called the 20S core and the regulator 19S formed by a 9-subunit base that binds to the ring of the 20S core, and a 10-subunit lid. PMSC1 is one of the 19 essential subunits of a 19S proteasome complex[59]. It is a 49-kDa protein comprised of 440 amino acids[60]. In 2012, Tsai *et al.*[1] have shown the interaction of PCDH10 with PDS-95 (**Figure 72**): PSD-95 is ubiquitinated by the ubiquitin E3 ligase murine double minute 2 (Mdm2) and then binds to Pcdh10, which links it to the proteasome for degradation. This degradation is responsible for synapse elimination. This study convinced us to include the candidate in the priority list to understand if it could play a role in this interaction since PMSC1 resulted to be candidate number 39 in the Array for PCDH10. Despite some inconclusive MAPPIT confirmation data, we included the candidate in further MAPPIT tests and CoIP experiments. The observation of this interaction could indicate a role of PCDHs in dendrite development and/or dendritic spine function: so far little evidence has been shown for this role (**Table 20**).

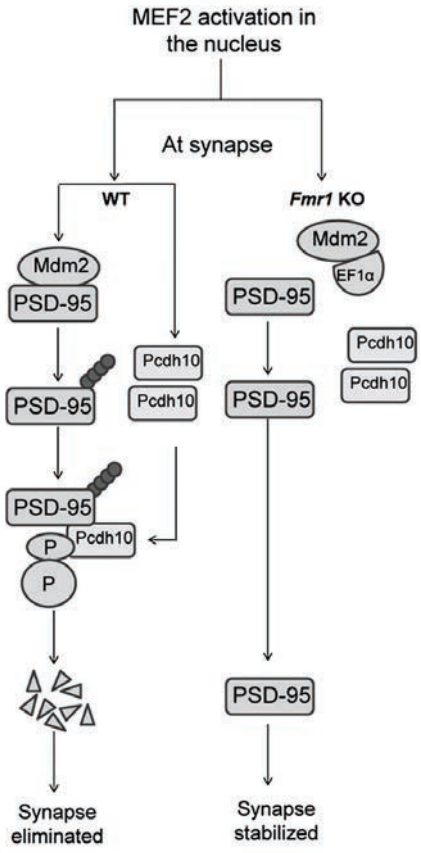


Figure 72 PCDH10 is required for synapse elimination.
Working model of MEF2-induced synapse elimination in WT neurons and the molecular basis of the deficit in synapse elimination in Fmr1 *KO neurons. Figure modified after*[1].

Table 20 δPcdh Dendritic Roles

Name	Reported Dendritic Role	Ref.
Protocadherin 7, NF-protocadherin (*Xenopus)*	Pcdh7 promotes retinal ganglion cell dendrite arborisation, possibly through TAF1.	[61]
Protocadherin 8, Arcadlin (mammals), paraxial	Pcdh8 causes a decrease in spine number by activation of a MAPK pathway resulting in Pcdh8/N-cadherin endocytosis. Consequently, neurons cultured from Pcdh8 knockout mice have increased dendritic spines, but only in the presence of N-cadherin.	[62]
Protocadherin 10, Ol-protocadherin	Pcdh10 knockdown prevented MEF2-induced synapse elimination. Pcdh10 is necessary for ubiquitination and proteasomal degradation of PSD-95, leading to spine loss.	[1]

IV.2.6 Actin-Related proteins

In a screen for novel PCDH10 interaction partners in lysates of newborn mouse brain, Nakao *et al.* identified NAP1 as a cytoskeleton-related protein[63]. Recently DYNLT1, also cytoskeleton-related, has been identified by MAPPIT and CoIP experiments as an interaction partner of PCDH10 and PCDH11X (I.Kahr and U. Fuchs, not published) and preliminary studies have also suggested an interaction with several other δPCDHs (our unpublished data).

The cytoskeleton is a very dynamic and adaptable network of cytoplasmic protein filaments and tubules, and is responsible for the proper cell functioning: it guarantees the correct cell shape, the ability to move and to organize the internal components and to interact mechanically with other cells and with the environment. It is responsible for the constitution of the mitotic division machinery, and in neural cells it is the support for dendrites and axon structure[64].

The cytoskeleton comprises three different kinds of proteic filaments: actin filaments or microfilaments, microtubules (formed by heterodimers of α and β tubulin) and a group of polymers called intermediate filaments[65]. The intermediate filaments show a number of different protein subunits, are the less dynamic filaments, are often assembled in response to mechanical stress and are involved in stability rather than movement. Microtubules instead are important mediators of cell movement, are the largest fibers and are formed of tubulin heterodimers which assemble to create

protofilaments, which in turn can assemble to form the microtubules. Microtubule-associated proteins or MAPs are proteins that bind to the microtubules and that can mediate the interaction between microtubules and other cell components. Motor proteins are proteins that can associate with the polarized fibers of cytoskeleton and that convert chemical energy in the form of ATP to mechanical energy. In case of microtubules there are two major classes of motor proteins: kinesins and dyneins. Kinesins are responsible for the fast anterograde axonal transport which is the movement of vesicles and organelles outward, from the cell body (also called soma) to the synapse or cell membrane while dyneins participate in the movement in the opposite direction, i.e. the retrograde axonal transport. The elaborate cytoskeletal structures of neurons are shown in **Figure 73**.

Dyneins are divided in cytoplasmic or axonemal. The cytoplasmic dyneins are responsible for the microtubule-mediated transport processes. They are large multisubunit complexes made of two heavy chains with the motor domain in combination with associated subunits such as intermediate (IC), light intermediate (LIC) and light chains (LC). Cytoplasmic dynein needs to associate with a large protein complex called dynactin in order to translocate organelles. The 74-kDa intermediate chain (IC74) mediates the binding while the other subunit is responsible for the direct binding to the cargo. DYNLT, DYNLL and DYNLRB are three distinct families of LCs. DYNLT1, one of the members of DYNLT family, has been shown to interact with several molecules and it has been suggested to play a role in the internalization of many cargos; in contrast, it has been shown to also have several dynein-independent functions. Our results show that PCDH10 interacts with DYNLT1, but not with the dynein-independent form, suggesting a role of the interaction in the motor complex (I. Kahr, not published). Our data indicate that PCDH10 might recruit DYNLT1 to the cell cortex for cell-cell migration or for initialization of the cell-cell contact process. Otherwise, it is possible that in different conditions PCDH10 is transported by DYNL1 to the perinuclear region, maybe for degradation. Anyway, the function of this interaction has not been clarified yet.

Finally, the third kind of protein filaments of the cytoskeleton is the microfilament. Microfilaments are formed of globular (G) actin: each monomer interacts head-to-tail with two other actin monomers, forming in this way filaments (filamentous (F) actin). The filaments of actin localize beneath the plasma membrane and are anchored to the plasma membrane at the adherents junctions. This contact is cadherin-mediated through the binding to catenins, able to associate with actin filaments.

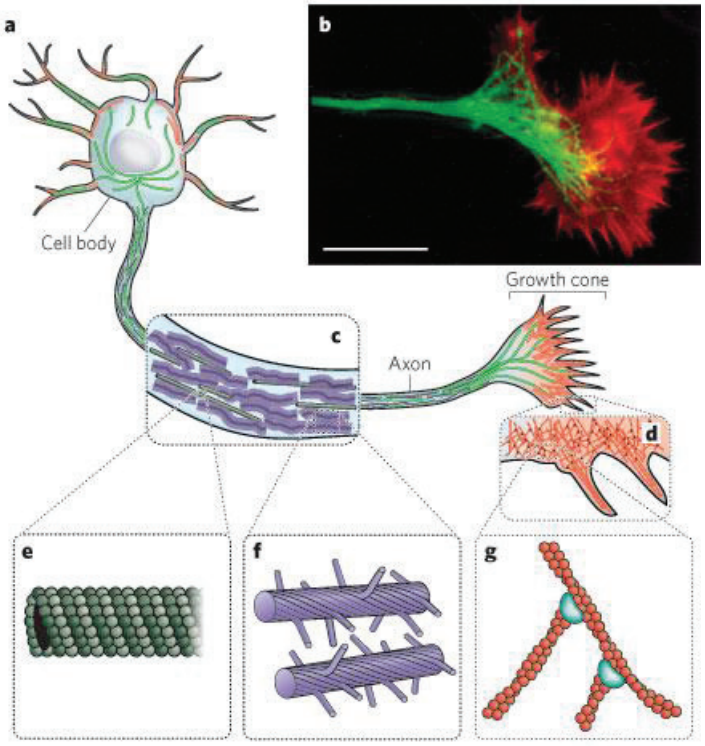


Figure 73 Neurons have elaborate cytoskeletal structures.
(A, B) Neurons are specialized eukaryotic cells that extend long processes to form connections in the nervous system. Like other eukaryotic cells, neurons have a cytoskeleton that consists of three main polymers: microtubules (green), intermediate filaments (purple) and actin filaments (red). Microtubules emanate from the axon, and actin-filament networks form sheet-like structures and filopodial protrusions at the leading edge. Scale bar, 20 µm. (C) The neuronal axon is a long membrane-bounded extension, in which neurofilaments (a class of intermediate filament in neurons) form a structural matrix that embeds microtubules, which transport materials from the cell body to the axon terminals at the synapse. (D) The growth cone contains dendritic actin filament networks and parallel actin filament filopodia. (E) Microtubules consist of 13 protofilaments of tubulin dimers arranged in a hollow tube. (F) Neurofilaments have flexible polymer arms that repel neighboring neurofilaments and determine the radius of the axon. (G) Actin filaments are arranged into networks. These networks can have many architectures, including the branched structures depicted here, which are formed by the Arp2/3 complex (blue). The diameters of microtubules, intermediate filaments, and actin filaments are within a factor of three of each other; the diagrams in E, F, and G are drawn approximately to scale. However, the relative flexibilities of these polymers differ markedly, as indicated by their persistence lengths: from least to most flexible, microtubules (5,000 µm), actin filaments (13.5 µm) and intermediate filaments (0.5 µm). Figure from[65].

Microfilaments, similar to microtubules, associate with motor proteins, in this case with myosin, responsible for contraction in the skeletal muscles. Networks of actin filaments constitute lamellipodia, protrusive actin structure localized at the leading edge of migrating cells, and filopodia, thin cytoplasmic projections that extend beyond the leading edge of lamellipodia[66,67]. Migrating cells show a fast turn-over of the actin cytoskeleton with lamellipodial protrusion boosted by actin polymerization. In cells, the major initiator of new actin filaments is the Arp2/3 complex[68]. Wiskott–Aldrich syndrome protein (WASP) and WASP-family verprolin-homologous protein (WAVE) family proteins are responsible for the activation of Arp2/3 complex. Wave protein is incorporated in a large complex (WAVE regulatory complex, WRC) together with PIR121, Nap1, Abi and HSPC300[69]. The WRC is activated by Rac, which directly binds to Sra1 causing the release of VCA, a conserved C-terminal sequence of the WASP protein responsible for the activation of Arp2/3[70]. Different (classes of) ligands have been identified for WRC activation, including PCDH10 and PCDH19. The conserved WAVE regulatory complex (WRC)-interacting receptor sequence (WIRS) mediates the interaction. Indeed, studies have demonstrated that PCDH10 and PCDH19 enhance activation of the WRC by Rac1 and consequently the production of new actin filaments[71].

PCDH10 was identified as a Nap1 interaction partner[72] and it was shown to pull down WAVE1 and ABI[63]. Nap1 (Nck associated protein 1) is an adaptor protein expressed in the cortical plate region of the developing cortex and it seems to play a role in the final steps of neuronal migration and differentiation[73]. Neurons have an elaborate cytoskeletal structures.

Studies conducted in astrocytoma have shown that PCDH10 can recruit NAP1 and WAVE1 to the cell-cell contact: no effect was noticed on the motility of single cells but the movement was accelerated in case of cell-cell contacts because of a reorganization of the assembly of F-actin and N-cadherin. Nakao and colleagues suggested a direct interaction of Nap1 and a region at the C-terminal of Pcdh10, downstream the CM2. This region was named Nap1-binding fragment (NBF). Recently the binding site of PCDH10 with WRC has been identified as a 9 residue sequence included in the NBF, which is necessary and sufficient for the binding. This motif has been named WIRS[74]. WIRS binds a surface of WRC composed by SRA1 and ABI2: this surface is created only when the complex is complete. These data are in contrast with a direct interaction of PCDH10 and Nap1 which seems to be indirect and related to a common partner such as Sri1[74]. Chicken Pcdh19 was shown to interact with Nap1 as well and the same study suggested an interaction also for Nap1 with Pcdh1 and Pcdh9[75].

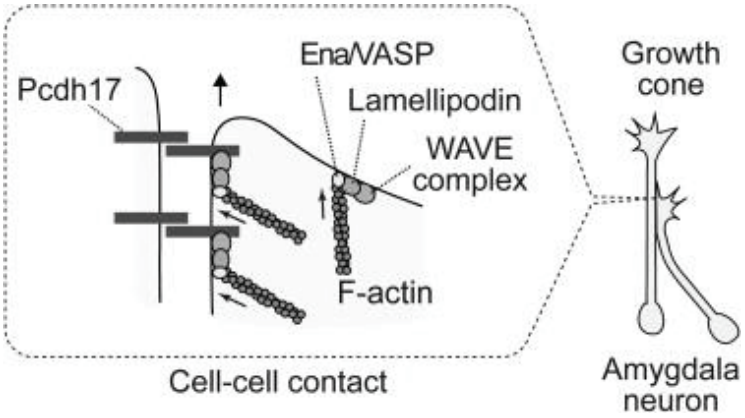


Figure 74 Pcdh17 recruits the WAVE complex, lamellipodin and Ena/Vas proteins at the cell-cell contact
Via recruiting of actin polymerization regulators to the contact site between neurons, Pcdh17 maintains the migration of the growth cones in neurons. The growth cones are in contact with other axons and are responsible for the fasciculation growth of axons from the same functional group of neurons. (Modified after[76]).

Recently, PCDH17 was shown to recruit actin regulator complexes to regulate a (selective) fasciculation in mammalian axons[76]. Using amygdala neurons of Pcdh17 KO mice, it has been demonstrated that Pcdh17 is required for normal axon extensions. Pcdh17 recruits the WAVE complex, lamellipodin and Ena/Vas proteins at the cell-cell contact.

If Pcdh17 forms homophilic interactions via its extracellular domain, the cytoplasmic domain seems to associate with the WAVE complex to create a motility machine and sustains cell migration process. In mouse brain lysate, Pcdh17 coimmunoprecipitates with Nap1, but also with Wave1 and Abi1. Taken together, Pcdh17 induces cell-cell contact by its extracellular domain and stimulates motility at the cell contact site by binding actin regulator molecules in the cytoplasm.

To summarize, δ2PCDHs have been shown to play a role in cell motility and to bind to WRC to influence actin cytoskeletal dynamics. WRC is involved in the regulation of actin polymerization, responsible for the movement of the cells. In view of this, we performed experiments and analyses focused on identification of novel relevant Pcdh interactions to elucidate how this complex mechanism is regulated.

IV.3 Conclusions

PCDH9, PCDH10 and PCDH11X were selected to investigate the PCDHs family. An Array MAPPIT screening was performed and results were confirmed in binary MAPPIT assay. After analysis and selection, a priority list was proposed to be further investigated. Coimmunoprecipitation has been used in this study as the prime confirmation test. In this way, I could confirm a molecular interaction of PCDH10 with MAX, MAD3, FHL2, CUL5, USP21 and USP46. For PCDH11X we confirmed the interaction with MAX, FHL2, USP21 and USP46. PDLIM7 was also considered a confirmed partner, despite problems with an aspecific signal in CoIP, as it is positive both in several MAPPIT screens and in a co-localization experiment.

The confirmed candidates belong to different pathways involved in many different biological mechanisms, proposing multiple interesting roles for PCDHs. We mainly focused on cancer related mechanisms to further investigate the identified novel interactions and their putative roles in cancer. For instance, candidates belonging to the MAX-MAD family (see chapter IV.2.1) play an important role in cell development and cell cycle. The c-Myc protein induces cell proliferation by targeting transcriptional expression of genes (including cyclin D1) leading to G0–G1 cell cycle progression. C-Myc promotes cell cycle progression through heterodimerization with MAX. If MAD proteins are present, MAX heterodimerizes with MAD and represses gene transcription: it associates with the mSin3 co-repressor complex via histone deacetylation[77]. The balance of the two MAX complexes is strictly regulated and disturbing events can lead to tumorigenesis and development, miss regulation of differentiation and cell growth. Since the pathway is responsible for gene regulation, any mutation might interfere with transcription. As described (IV.2.1), our results indicated MAX, MXD3 and MXD1 as putative interaction partners of different members of the δPCDH family. The candidates are mutated in brain cancer and are defined as tumor suppressor genes.

Also CUL5 (section IV.2.5), which mediates ubiquitination and subsequent proteasomal degradation of targeted proteins, is a tumor suppressor gene that inhibits cell proliferation. Together with PMSC1 and USPs proteins (section IV.2.4), CUL5 belongs to the ubiquitination pathway responsible for protein degradation. The ubiquitin-proteasome system is responsible for the regulation of cyclin-dependent kinases (CDKs) and of other cell cycle components but also for other pathways that interfere with the cell cycle program such as stress responses or checkpoint systems. This demonstrates the enormous impact of the ubiquitin-proteasome system on cell proliferation. Our

investigation identified various member of this pathway as PCDHs interaction partners. PCDH10 is required for synapse elimination and this is an interesting indication of the role that PCDHs might play in the degradation system: the mechanism has been not completely elucidated yet. In this context PMSC1 was a top candidate: we suggested that it could be responsible for the connection between PCDH10 and the proteasome. Unfortunately, it was not possible to confirm this interaction in CoIPs and further investigations with different techniques should be performed to better investigate this putative interaction.

Migration and proliferation events are regulated also by FHL2/3, which are additional candidates on the priority list (section IV.2.2). FHL2/3 is known to be an important inducer of the EMT process. Our CoIP experiments confirmed the interaction of FHL2 with both PCDH10 and PCDH11X, suggesting again a role of different members of the δPCDH family in cell proliferation. Furthermore, δ2PCDHs have been shown to play a role in the reorganization of actin cytoskeleton, as occurring in migrating cells (section IV.2.6). By MAPPIT technology, our group identified DYNLT1 as an interaction partner of PCDH10 and PCDH11X and confirmed this by CoIP. The MAPPIT array indicated a list of novel candidate interactors which are all actin-cytoskeleton related. Considering the indications concerning cell migration being influenced by δPCDHs, as obtained by our and published studies, this class of molecules deserves to be further studied in relation to the δPCDH family members.

To summarize, the selected candidate interactors are cancer-related (and/or play a role in neurological diseases), as they mostly regulate cell cycle, proliferation and transcription. Different evidence indicates that δPCDHs are involved in tumorigenesis and tumor progression and are mainly expressed in brain: in view of this, the confirmed proteins of the candidate list of interactors appear to be promising and interesting candidates. Since δPCDH functions are largely unknown, some functional experiments were performed on the basis of the reported properties and functions of the interaction candidates (described in Chapter V).

IV.4 References

1. Tsai, N.-P. *et al.* Multiple autism-linked genes mediate synapse elimination via proteasomal degradation of a synaptic scaffold PSD-95. *Cell* **151,** 1581–94 (2012).

2. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* **16,** 653–99 (2000).

3. Lüscher, B. Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene* **277,** 1–14 (2001).

4. Luscher, B. & Eisenman, R. N. New light on Myc and Myb. Part I. Myc. *Genes Dev.* **4,** 2025–2035 (1990).

5. Blackwood, E. M., Luscher, B. & Eisenman, R. N. Myc and Max associate in vivo. *Genes Dev.* **6,** 71–80 (1992).

6. Zhang, H., Fan, S. & Prochownik, E. V. Distinct roles for MAX protein isoforms in proliferation and apoptosis. *J. Biol. Chem.* **272,** 17416–24 (1997).

7. Ayer, D. E. & Eisenman, R. N. A switch from Myc:Max to Mad:Max heterocomplexes accompanies monocyte/macrophage differentiation. *Genes Dev.* **7,** 2110–9 (1993).

8. Zervos, A. S., Gyuris, J. & Brent, R. Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell* **72,** 223–32 (1993).

9. Hurlin, P. J. *et al.* Mad3 and Mad4: novel Max-interacting transcriptional repressors that suppress c-myc dependent transformation and are expressed during neural and epidermal differentiation. *EMBO J.* **14,** 5646–59 (1995).

10. Comino-Méndez, I. *et al.* Exome sequencing identifies MAX mutations as a cause of hereditary pheochromocytoma. *Nat. Genet.* **43,** 663–7 (2011).

11. Cascón, A. & Robledo, M. MAX and MYC: a heritable breakup. *Cancer Res.* **72,** 3119–24 (2012).

12. Kohl, N. E. *et al.* Transposition and amplification of oncogene-related sequences in human neuroblastomas. *Cell* **35,** 359–67 (1983).

13. O'Hagan, R. C. *et al.* Gene-target recognition among members of the myc superfamily and implications for oncogenesis. *Nat. Genet.* **24,** 113–9 (2000).

14. Ogawa, H., Ishiguro, K.-I., Gaubatz, S., Livingston, D. M. & Nakatani, Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296,** 1132–6 (2002).

15. Wechsler, D. S., Shelly, C. A., Petroff, C. A. & Dang, C. V. MXI1, a putative tumor suppressor gene, suppresses growth of human glioblastoma cells. *Cancer Res.* **57,** 4905–12 (1997).

16. Eagle, L. R. *et al.* Mutation of the MXI1 gene in prostate cancer. *Nat. Genet.* **9,** 249–55 (1995).

17. Yun, J.-S., Rust, J. M., Ishimaru, T. & Díaz, E. A novel role of the Mad family member Mad3 in cerebellar granule neuron precursor proliferation. *Mol. Cell. Biol.* **27,** 8178–89 (2007).

18. Barisone, G. A., Yun, J.-S. & Díaz, E. From cerebellar proliferation to tumorigenesis: new insights into the role of Mad3. *Cell Cycle* **7,** 423–7 (2008).

19. Morgan, M. J. & Madgwick, A. J. The LIM proteins FHL1 and FHL3 are expressed differently in skeletal muscle. *Biochem. Biophys. Res. Commun.* **255,** 245–50 (1999).

20. Kadrmas, J. L. & Beckerle, M. C. The LIM domain: from the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* **5,** 920–31 (2004).

21. Johannessen, M., Møller, S., Hansen, T., Moens, U. & Van Ghelue, M. The multifunctional roles of the four-and-a-half-LIM only protein FHL2. *Cell. Mol. Life Sci.* **63,** 268–84 (2006).

22. Kurakula, K., van der Wal, E., Geerts, D., van Tiel, C. M. & de Vries, C. J. M. FHL2 protein is a novel co-repressor of nuclear receptor Nur77. *J. Biol. Chem.* **286,** 44336–43 (2011).

23. Fimia, G. M., De Cesare, D. & Sassone-Corsi, P. A family of LIM-only transcriptional coactivators: tissue-specific expression and selective activation of CREB and CREM. *Mol. Cell. Biol.* **20,** 8613–22 (2000).

24. Li, H. Y. *et al.* Protein-protein interaction of FHL3 with FHL2 and visualization of their interaction by green fluorescent proteins (GFP) two-fusion fluorescence resonance energy transfer (FRET). *J. Cell. Biochem.* **80,** 293–303 (2001).

25. Niu, C. *et al.* Downregulation and antiproliferative role of FHL3 in breast cancer. *IUBMB Life* **63,** 764–71 (2011).

26. Gabriel, B. *et al.* Expression of the transcriptional coregulator FHL2 in human breast cancer: a clinicopathologic study. *J. Soc. Gynecol. Investig.* **13,** 69–75 (2006).

27. Gabriel, B. *et al.* Focal adhesion kinase interacts with the transcriptional coactivator FHL2 and both are overexpressed in epithelial ovarian cancer. *Anticancer Res.* **24,** 921–7

28. Wang, J. *et al.* Suppression of FHL2 expression induces cell differentiation and inhibits gastric and colon carcinogenesis. *Gastroenterology* **132,** 1066–76 (2007).

29. Arimura, T. *et al.* Structural analysis of four and half LIM protein-2 in dilated cardiomyopathy. *Biochem. Biophys. Res. Commun.* **357,** 162–7 (2007).

30. Kinoshita, M., Nakagawa, T., Shimizu, A. & Katsuoka, Y. Differently regulated androgen receptor transcriptional complex in prostate cancer compared with normal prostate. *Int. J. Urol.* **12,** 390–7 (2005).

31. Müller, J. M. *et al.* The transcriptional coactivator FHL2 transmits Rho signals from the cell membrane into the nucleus. *EMBO J.* **21,** 736–48 (2002).

32. Yan, Q. *et al.* KLF8 promotes tumorigenesis, invasion and metastasis of colorectal cancer cells by transcriptional activation of FHL2. *Oncotarget* **6,** 25402–17 (2015).

33. Al-Nomani, L., Friedrichs, J., Schüle, R., Büttner, R. & Friedrichs, N. Tumoral expression of nuclear cofactor FHL2 is associated with lymphatic metastasis in sporadic but not in HNPCC-associated colorectal cancer. *Pathol. Res. Pract.* **211,** 171–4 (2015).

34. Verset, L. *et al.* Epithelial expression of FHL2 is negatively associated with metastasis-free and overall survival in colorectal cancer. *Br. J. Cancer* **109,** 114–120 (2013).

35. Verset, L., Feys, L., Trépant, A.-L., De Wever, O. & Demetter, P. FHL2: a scaffold protein of carcinogenesis, tumour-stroma interactions and treatment response. *Histol. Histopathol.* **31,** 469–78 (2016).

36. Zhang, W. *et al.* Four and a half LIM protein 2 (FHL2) negatively regulates the transcription of E-cadherin through interaction with Snail1. *Eur. J. Cancer* **47,** 121–30 (2011).

37. Miyamoto, K. *et al.* Identification of 20 genes aberrantly methylated in human breast cancers. *Int. J. Cancer* **116,** 407–14 (2005).

38. Ying, J. *et al.* Functional epigenetics identifies a protocadherin PCDH10 as a candidate tumor suppressor for nasopharyngeal, esophageal and multiple other carcinomas with frequent methylation. *Oncogene* **25,** 1070–80 (2006).

39. Ying, J. *et al.* Frequent epigenetic silencing of protocadherin 10 by methylation in multiple haematologic malignancies. *Br. J. Haematol.* **136,** 829–32 (2007).

40. Yu, J. *et al.* Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* **136,** 640–51.e1 (2009).

41. Kleiber, K., Strebhardt, K. & Martin, B. T. The biological relevance of FHL2 in tumour cells and its role as a putative cancer target. *Anticancer Res.* **27,** 55–61

42. Fanning, A. S. & Anderson, J. M. Protein-protein interactions: PDZ domain networks. *Curr. Biol.* **6,** 1385–8 (1996).

43. Kim, S. H., Yamamoto, A., Bouwmeester, T., Agius, E. & Robertis, E. M. The role of paraxial protocadherin in selective adhesion and cell movements of the mesoderm during Xenopus gastrulation. *Development* **125,** 4681–4690 (1998).

44. Kornau, H. C., Schenker, L. T., Kennedy, M. B. & Seeburg, P. H. Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. *Science* **269,** 1737–40 (1995).

45. Songyang, Z. Recognition of Unique Carboxyl-Terminal Motifs by Distinct PDZ Domains. *Science (80-.).* **275,** 73–77 (1997).

46. Borrello, M. G. *et al.* Differential interaction of Enigma protein with the two RET isoforms. *Biochem. Biophys. Res. Commun.* **296,** 515–22 (2002).

47. Jung, C.-R. *et al.* Enigma negatively regulates p53 through MDM2 and promotes tumor cell survival in mice. *J. Clin. Invest.* **120,** 4493–506 (2010).

48. Durick, K., Gill, G. N. & Taylor, S. S. Shc and Enigma are both required for mitogenic signaling by Ret/ptc2. *Mol. Cell. Biol.* **18,** 2298–308 (1998).

49. Schalm, S. S., Ballif, B. A., Buchanan, S. M., Phillips, G. R. & Maniatis, T. Phosphorylation of protocadherin proteins by the receptor tyrosine kinase Ret. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13894–9 (2010).

50. Phillips, G. R. *et al.* Gamma-protocadherins are targeted to subsets of synapses and intracellular organelles in neurons. *J Neurosci* **23,** 5096–5104 (2003).

51. Blank, M., Triana-Baltzer, G. B., Richards, C. S. & Berg, D. K. Alpha-protocadherins are

presynaptic and axonal in nicotinic pathways. *Mol Cell Neurosci* **26,** 530–543 (2004).

52. Komander, D. The emerging complexity of protein ubiquitination. *Biochem. Soc. Trans.* **37,** 937–53 (2009).

53. Daviet, L. & Colland, F. Targeting ubiquitin specific proteases for drug discovery. *Biochimie* **90,** 270–83 (2008).

54. Vucic, D., Dixit, V. M. & Wertz, I. E. Ubiquitylation in apoptosis: a post-translational modification at the edge of life and death. *Nat. Rev. Mol. Cell Biol.* **12,** 439–52 (2011).

55. Lecker, S. H., Goldberg, A. L. & Mitch, W. E. Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J. Am. Soc. Nephrol.* **17,** 1807–19 (2006).

56. Nijman, S. M. B. *et al.* A genomic and functional inventory of deubiquitinating enzymes. *Cell* **123,** 773–86 (2005).

57. USP gene family. (2015). at <http://ghr.nlm.nih.gov/geneFamily/usp>

58. North, W. G., Fay, M. J., Longo, K. A. & Du, J. Expression of all known vasopressin receptor subtypes by small cell tumors implies a multifaceted role for this neuropeptide. *Cancer Res.* **58,** 1866–71 (1998).

59. Gu, Z. C. & Enenkel, C. Proteasome assembly. *Cell. Mol. Life Sci.* **71,** 4729–45 (2014).

60. http://www.uniprot.org/uniprot/P62191 . at <http://www.uniprot.org/uniprot/P62191>

61. Piper, M., Dwivedy, A., Leung, L., Bradley, R. S. & Holt, C. E. NF-protocadherin and TAF1 regulate retinal axon initiation and elongation in vivo. *J. Neurosci.* **28,** 100–5 (2008).

62. Yasuda, S. *et al.* Activity-induced protocadherin arcadlin regulates dendritic spine number by triggering N-cadherin endocytosis via TAO2beta and p38 MAP kinases. *Neuron* **56,** 456–71 (2007).

63. Nakao, S., Platek, A., Hirano, S. & Takeichi, M. Contact-dependent promotion of cell migration by the OL-protocadherin-Nap1 interaction. *J. Cell Biol.* **182,** 395–410 (2008).

64. Alberts, B. *et al.* Molecular Biology of the Cell. (2002).

65. Fletcher, D. A. & Mullins, R. D. Cell mechanics and the cytoskeleton. *Nature* **463,** 485–92 (2010).

66. Mattila, P. K. & Lappalainen, P. Filopodia: molecular architecture and cellular functions. *Nat. Rev. Mol. Cell Biol.* **9,** 446–454 (2008).

67. Machesky, L. M. Lamellipodia and filopodia in metastasis and invasion. *FEBS Lett.* **582,** 2102–2111 (2008).

68. Krause, M. & Gautreau, A. Steering cell migration: lamellipodium dynamics and the regulation of directional persistence. *Nat. Rev. Mol. Cell Biol.* **15,** 577–90 (2014).

69. Eden, S., Rohatgi, R., Podtelejnikov, A. V, Mann, M. & Kirschner, M. W. Mechanism of regulation of WAVE1-induced actin nucleation by Rac1 and Nck. *Nature* **418,** 790–3 (2002).

70. Chen, Z. *et al.* Structure and control of the actin regulatory WAVE complex. *Nature* **468,** 533–8 (2010).

71. Chen, B. *et al.* The WAVE Regulatory Complex Links Diverse Receptors to the Actin Cytoskeleton. *Cell* **156,** 195–207 (2014).

72. Uemura, M., Nakao, S., Suzuki, S. T., Takeichi, M. & Hirano, S. OL-Protocadherin is essential for growth of striatal axons and thalamocortical projections. *Nat. Neurosci.* **10,** 1151–9 (2007).

73. Yokota, Y., Ring, C., Cheung, R., Pevny, L. & Anton, E. S. Nap1-regulated neuronal cytoskeletal dynamics is essential for the final differentiation of neurons in cerebral cortex. *Neuron* **54,** 429–45 (2007).

74. Chen, B. *et al.* The WAVE regulatory complex links diverse receptors to the actin cytoskeleton. *Cell* **156,** 195–207 (2014).

75. Tai, K., Kubota, M., Shiono, K., Tokutsu, H. & Suzuki, S. T. Adhesion properties and retinofugal expression of chicken protocadherin-19. *Brain Res.* **1344,** 13–24 (2010).

76. Hayashi, S. *et al.* Protocadherin-17 Mediates Collective Axon Extension by Recruiting Actin Regulator Complexes to Interaxonal Contacts. *Dev. Cell* **30,** 673–687 (2014).

77. Schreiber-Agus, N. *et al.* Drosophila Myc is oncogenic in mammalian cells and plays a role in the diminutive phenotype. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 1235–40 (1997).

78. Aravind, L. & Koonin, E. V. The U box is a modified RING finger — a common domain in ubiquitination. *Curr. Biol.* **10,** R132–R134 (2000).

Chapter V. FUNCTIONAL STUDIES [f]

[f]The study was designed by Eleonora Billi and Frans van Roy.

Pcdh10all and Pcdh10long KO mice were generated by Irene Kahr.

All experiments were done by Eleonora Billi unless stated otherwise.

RNA seq experiments were performed by the VIB Nucleomics core.

RNA Seq analyses were done by Eleonora Billi, Pieter De Bleser and Frans van Roy.

Quantitative analyses of cell properties of cell lines with induced protocadherin isoform expression were performed by Eleonora Billi and Marleen van Troys.

V.1 Introduction

Despite the steadily growing interest in investigating the functions of δPCDH family members, mostly due to their possible clinical impact, many details still remain obscure. Pcdhs are probably communicative rather than adhesive proteins and functional differences may exist between long and short isoforms. In this study we first identified and confirmed different candidates for δPCDHs, and then tried to investigate the role of the confirmed interactions. It was difficult to determine *ex ante* how to proceed best in terms of functional characterization. Since the functions of δPCDHs are largely unknown, functional characterization was much dependent on the identity of the candidates. Functional implications of the observed molecular interactions for normal morphogenesis as well as for tumorigenesis and tumor progression were the targets of the investigation with contemporary molecular tools.

Different candidate interactors were analyzed with different techniques according to the known specific functions they might play in health or disease, and we focused on their putative or proven roles in the central neural system and cancer. First, we studied expression levels for the candidate interactors by use of Pcdh10 KO mice, and then we proceeded with specific tests for protein ubiquitination, protein half-life and degradation, cell migration, proliferation and adhesion.

This Chapter thus covers three different approaches to further investigate the functional indications obtained from the identification of novel interaction partners.

In the first part, a comparison of the expression level between PCDH10 KO mice and WT mice is presented: RNA sequencing experiments for cerebellum and qRT-PCR tests for total brain transcripts of PCDH10 KO versus WT mice will be presented and analyzed.

In the second part specific validation studies for novel PCDH interaction partners are described including effects of selected interaction partners on PCDH ubiquitination and effect of PCDH expression on Myc activity.

In the third and last part of the results section of this dissertation, a comparative functional analysis of specific cellular properties of PCDH10 and PCDH11X expressing cell lines will be presented. We established an in vitro cell based assay with the aim to further investigate the role of different PPIs on these selected properties.

V.2 Expression Analysis

V.2.1 Background

Surprisingly, most of the candidate interaction partners of our priority list for different δPCDHs were transcription factors. The first test we therefore decided to perform was gene expression analysis. Experiments were carried out in mouse brain RNA samples.

To quantify RNA molecules, the classic technique has been the Northern blot, since 1977: the technique reveals RNA expression using molecular probes and uses the principle of nucleic acid hybridization[1]. With the time it has been defined an inconvenient technique since it is not very precise, it requires long experimental time and a large amount of RNA, a requirement which is not always easy to achieve. With the advent of RT-PCR (Reverse transcription polymerase chain reaction) thanks to the discovery of reverse transcriptase and PCR technology, the quantification of RNA found a new method of choice.

We performed qRT-PCR (Reverse transcription quantitative real-time PCR) experiments on brain RNA samples from Pcdh10 WT and KO mice, defective for either all isoforms or only the long isoforms. The expression levels of the candidates and related genes have been studied. Gene expression was also studied by RNA sequencing (RNA Seq) experiments.

V.2.2 Methodology

V.2.2.1 PCDH10 KO mice and genotyping

Conditional Pcdh10 knockout mice were generated in the lab by Irene Kahr using the recombineering technology (Summarized in **Figure 75**).

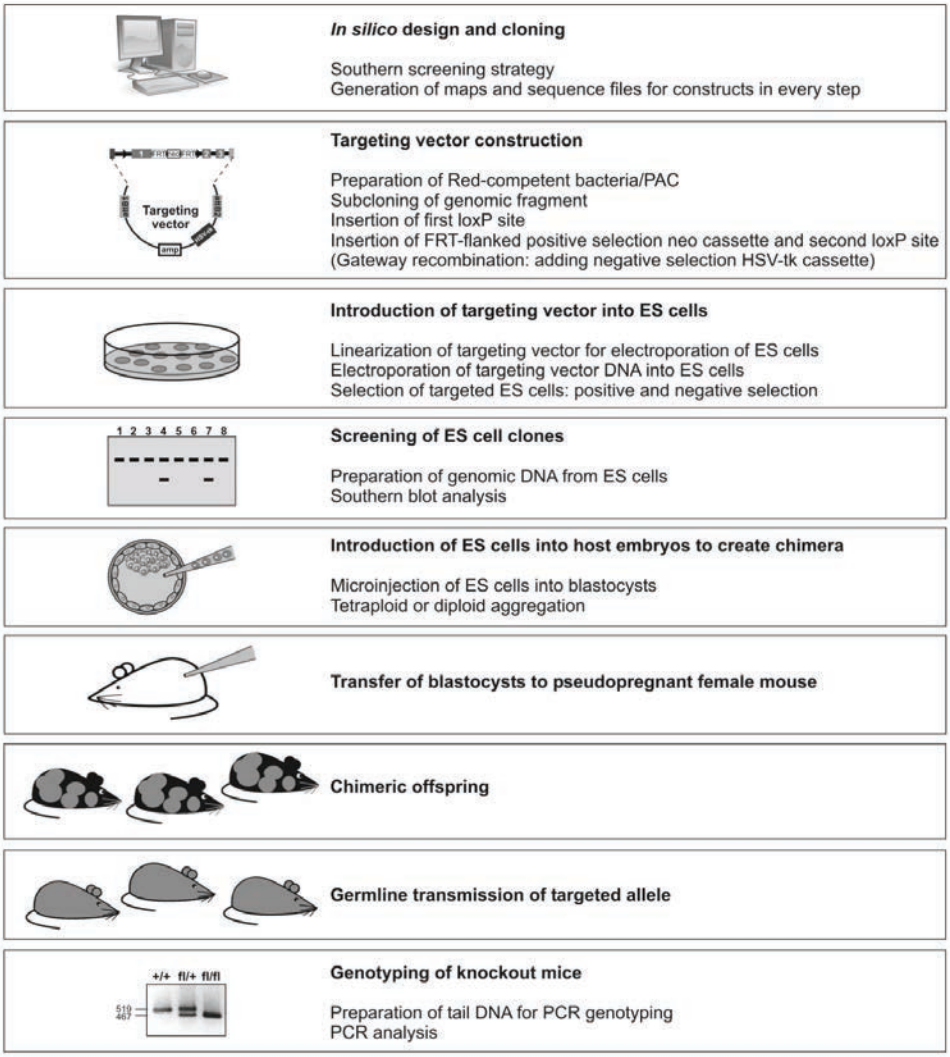


Figure 75 Outline of generation process of KO mice
Mice generation and picture by Kahr I. not published

Two different Pcdh10 mouse models were generated: Pcdh10all KO mice, where all the isoforms are deleted, and Pcdh10long KO mice where the short isoform (lacking the CM motifs) is conserved. All mice models were hosted in the SPF facility of the Inflammation Research Centre (IRC-VIB, UGent). Breeding and experiments were performed in the SFP facility according to the regulation and guidelines of the Ethical Committee for care and use of laboratory animals of the University of Gent.

Figure 76 shows the gene structure and the alternative transcripts for PCDH10/Pcdh10. Details on genomic organization of PCDH10 and of the other members of the family are described in paragraph I.3.1 of this dissertation.

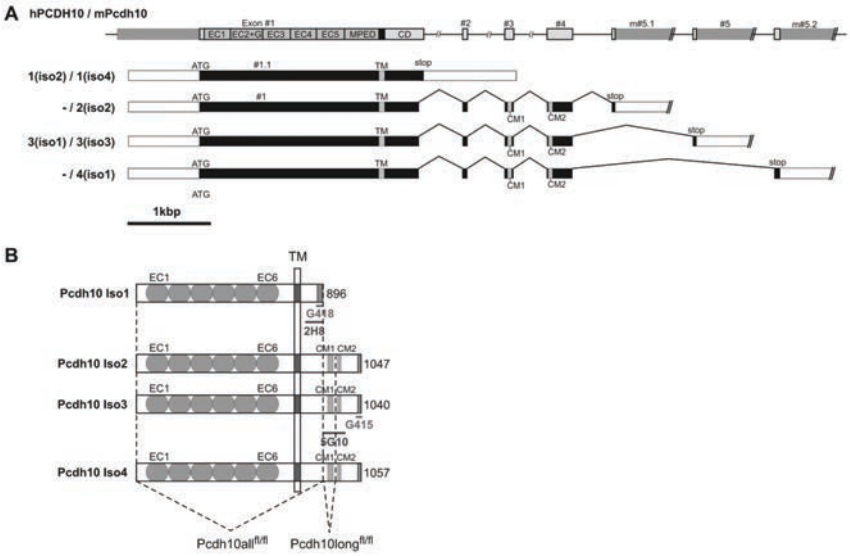


Figure 76 Gene structure and alternative transcripts of PCDH10/pcdh10 and Pcdh10 isoforms protein structure

(A) In the gene: grey boxes represent exons, the intervening line represent introns. In the transcripts: black boxes are the coding sequences, white boxes represent 5' or 3' untranslated regions. CD, sequence encoding the membrane proximal part of the cytoplasmic domain; CM1 and CM2 are the conserved motifs; EC, region encoding the cadherin extracellular repeats; MPED, sequence encoding the membrane proximal extracellular domain; TM encodes the transmembrane domain. Exon 1 alone encodes the complete ectodomain, the TM domain and a large part of the cytoplasmic domain. If no splicing occurs in the transcript a short isoform lacking the CMs is produced. In case of long isoforms, a splice donor is used in exon #1, and a long transcript is formed comprising exons encoding CM1 and CM2. (B) Schematic representation of Pcdh10 protein isoforms. Numbers on the right indicate the total size (number of amino acid residues) of each isoform. Dashed lines indicate the regions that were floxed in Pcdh10all fl/fl and Pcdh10long fl/fl mice, respectively. After the Cre mediated deletion of exon 1 as expected a complete Pcdh10 KO mice was obtained, while in the Pcdh10long fl/fl mice, the removal of exon 2 and 3 leads to a frame shift in exon 4, and therefore only the short isoform can be expressed. (Modified from Kahr I., not published)

Pcdh10all KO mice and Pcdh10long KO mice were genotyped using genomic DNA extracted from the tail. Tail tips were incubated overnight (O/N) at 55°C in 200 µl of lysis buffer (GB 10x, 10% Triton x-100, H_2O), (stock 200 ml GB 10x stored at -20°C : 134ml 2M Tris pH 8.8, 33,2 ml 1M $(NH_4)_2SO_4$, 21.4 ml 0.5 M $MgCl_2$ and 11,4 ml H_2O) and 0.8 µl proteinase K (10 mg/ml). Samples were then incubated for 10 min at 95°C. The genomic DNA was precipitated with isopropanol, washed with 70% ethanol, air-dried and dissolved in 300 µl of H_2O. The PCR reaction included: 0.2 µl of each specific primer (25 pmol), 2 µl of template, 0.4 µl dNTPs (10 mM), 2 µl of 10X PCR buffer, 1µl $MgCl_2$ (50 mM) 0.2 µl Taq and 14 µl H_2O. Used primers were for Floxed allele: Pcdh10 all 5' Forward CATGTACCTTCTTCCCACAC; Reverse GGCATGTGTCAATCAAAGCT; product size of wild type allele 519 bp, of targeted allele 469 bp. Pcdh10 Long 3' Forward CTTCCATTGGTCACTGTGCT; Reverse CCACCTTTGGCCATTAGTTA; product size of wild type allele 330 bp, of targeted allele 428 bp. For Cre-mediated deleted alleles: Pcdh10 all Forward CATGTACCTTCTTCCCACAC, Reverse GGTCTCTCAACTAGATAGCT, product size 775 bp. Pcdh10 long Forward GCGTGTACCAGTAAGCAATA, Reverse CCACCTTTGGCCATTAGTTA, product size 652 bp. The thermocycling conditions for PCR reactions forPcdh10 all were: 3 min at 94°C , 35 cycles of 94°C for 30 sec, 57°C for 30 sec and 72°C for 45 sec (1 min for pcdh10 long); 72°C for 10 min.

Mice were also genotyped to confirm the gender, by performing PCRs for specific regions only found in the Y chromosome: *DDX3Y* and *Eif2S3Y* genes were used for this purpose. Primers: DDX3Y Forward TGTTAGTTGCCACACCAGGACG, Reverse TGGTGGCATTGTGTCCTGCTCA; Eif2S3Y Forward GTATGCTGCTCCAGGTGGTCTT, Reverse CTCAGGTAATGCTCCAACAGCAC. Sry was used as well to confirm: Forward AACAACTGGGCTTTGCACATTG , Reverse GTTTATCAGGGTTTCTCTCTAGC.

V.2.2.2 Tissues and RNA extraction

Brains were isolated from WT, Pcdh10all and Pcdh10long homo and heterozygous KO mice. All mice were 3 weeks old. Brains were dissected in 3 parts: the olfactory bulbs, the cerebellum (separated from the colliculus and the spinal cord) and the rest. Most of the experiments were performed in cerebellum because of the high expression of Pcdh10.

Dissected brains were quick frozen and conserved at -80°C.

Total RNA was extracted in RNAse free conditions using the Aurum Total RNA mini Kit (Bio-Rad) according to the manufacturer's protocol. TRIzol (800 µl) was used for homogenation of the fatty brain tissue .160µl RNase free Chloroform were added to each sample: the Eppendorf tubes were inverted for 15 sec, incubated for 2-3 minutes and then centrifuged at 12000X g at 4°C for 10

minutes. The supernatant was transferred into a new tube, paying attention not to touch the interphase of proteins, and 60% RNase free ethanol was added to the supernatant in a 1:1 ratio. Columns from the Aurum Total RNA mini Kit were used and the sample was loaded in more times in a max amount of 600 µl each time. The protocol from the company was then followed. At the end of the process, the eluate is the total RNA sample and this was used immediately for downstream applications or frozen for later use. Cerebellum was divided in two equal parts and used separately. Elution was done in 80 µl of elution buffer from the kit. Sample concentrations and purity were checked with a Thermo Scientific NanoDrop™ Spectrophotometer. We used samples of cerebellum which were showing a concentration higher than 350 ng/ µl and samples of the cerebrum (without cerebellum) higher than 1200 ng/ul. The RNA purity was determined as a 260/280 wavelength ratio; a pure sample has a ratio of 2.1. The ratio 260/230 pointed at contaminations of salt or rest of phenol or presence of proteins. We considered clean sample with the ratio equal or higher than 2. Samples not meeting those characteristics were not used.

V.2.2.3 QRT-PCR

cDNA was prepared from brain RNA with the iScript cDNA synthesis kit according to the manufacturer's instructions (Bio-Rad). Quantitative RT-PCR mixes contained 20 ng template cDNA, Light Cycler 480 SYBR Green I Mastermix (Roche) and 300 nM forward and reverse primers. RT-PCR mixes for the 3 most stable reference genes, selected from a panel of reference genes by applying geNorm[2], were prepared in the same way. Reactions were performed on a Light Cycler 480 (Roche) using the following protocol: incubation at 95°C for 5 min, then 50 cycles at 95°C for 10 sec, 60°C for 30 sec, and 72°C for 1 sec. Specific primers were selected based on specificity (melting curve analysis) and efficiency (standard curve). List of primers: **Table 21**.

Table 21 Quantitative RT-PCR primers

Gene Name	Forward primer sequence	Reverse primer sequence
5330417C22Rik	CCTGAGTCTCTGTGGAAACCAG	ACAACCTGGCAGACATAGGCTG
Abl1	TGAGCAGAAAGATGCGCCTGAC	CGCTCATCTTCATTTAGGCTGCC
Adcy1	CCTTCTCCAACGTGATGACCTG	GCTGATGTACCTGTTGACGCGT
Basp1	CGAAGCGAGCGCCGAGAGCA	GCTAGGTTTAGAGTCTGAAGCCG
Bax	TCAGGATGCGTCCACCAAGAAG	TGTGTCCACGGCGGCAATCATC
Bid	CCACAACATTGCCAGACATCTCG	TCACCTCATCAAGGGCTTTGGC
Bmp3	TAACACGGTCCGCAGCTTCAGA	TGTGGCTGACAAAATGTTCTCCG
Brca1	CGAGGAAATGGCAACTTGCCTAG	TCACTCTGCGAGCAGTCTTCAG
Cartpt	TACTCTGCCGTGGATGATGCGT	TCGGAATGCGTTTACTCTTGAGC

Casp 1	GGCACATTTCCAGGACTGACTG	GCAAGACGTGTACGAGTGGTTG
Casp8	AGAAGAGGGTCATCCTGGGAGA	TCAGGACTTCCTTCAAGGCTGC
Cbln3	AGTGGTGCCATCTACTTCGACC	CCTGGACAGTTTGGCGGTTGTA
Ccnb1	AGAGGTGGAACTTGCTGAGCCT	GCACATCCAGATGTTTCCATCGG
Ccnd1	TCTACACCGACAACTCCATCCG	TCTGGCATTTTGGAGAGGAAGTG
Cdc2L1	GACTCTCCAGCACTGTCACCTA	GTCCTTTGCTCTGTAGACCACC
Cdh23	CACTGTCACCTATGCCATCGTG	GGCAGTGACAATCAGGGTGTAG
Cdon	TGAAGGACAGCCTGCCATGCTT	CCTGGAGGAATCCGTAAGCACA
Ceb1	CTGTGACTATCAAGACCATGCTG	AGGAGGGTCATCTTTGTGGTGG
Cerkl	CCTCTTGGCTTCATTCCAGCAG	GCAGACATCCACTGACCGTATG
Cetn4	GATGACGATGCTACGGGAAGCA	CCGTCACGATCAGCTTCATCAAG
C-Myc	TGAGCCCCTAGTGCTGCAT	AGCCCGACTCCGACCTCTT
Cst3	CATACAGGTGGTGAGAGCTCGT	CTGGTCATGGAAAGGACAGTCAG
Cul5	GTGTAAATGCGCTGGTGACCTC	CTCCAAGTCCTTCAACATCGGC
Cux1	CCAAGAACGGCATTTGCCAGAG	CCTTTCTGGGTCAGTTTGCTCC
Dact2	GCCTTGGATAGTGACAGCAGAC	GGATAGACGGTCGCTGCAAACA
Dkk3	GGAGGAAGCTACGCTCAATGAG	TTGCCAGGTTCACCTCAGAGGA
Dos	GGTCAAGAAATGGAAGCTGGAGC	TCCTGCTCCATGCTCTCACTGG
Dusp1	CAACCACAAGGCAGACATCAGC	GTAAGCAAGGCAGATGGTGGCT
Eed	AGTTGAGCAGCGACGAGAACAG	TCCTTCCTGGTGCATTTGGCGT
Ephb6	GGAACATCCTCGACTACCAGCT	TCAACCGTGTCACAGTGGCAGT
Exosc9	TCTCAGATGGCTGCTCCAGCTT	CCTTTTCACCAGCGACAACACAG
F2R	TTTCCTCTCCGACAGTCCTGGT	CAGCACAAGATGCTGTAGAGGTG
Fas	GGACCCAGAATACCAAGTGCAG	GTTGCTGGTGAGTGTGCATTCC
Fbl	GCAAGAGAGGAAACCAGTCAGG	AGACTCTCCAGGGACCAGATTC
Fhl1	GCTGCCTGAAGTGCTTTGACAAG	ACTTGGCACAGCGGAAGCAGTT
Fhl2	CCTGCTTCTGTGACTTGTACGC	CAGTCGTTGTGCCATTGGCGTT
Fhl3	ACCTGTCAGGACAGTGAGCTTC	GCACAGAAAGCAGTGTTCATGCC
Fndc9	CACGAACCTCGGTGGTCTTACA	TCTTTCGTGGCAGTTCCTCCAG
Fxyd7	CTATGCCACTGTGCAGACTGTG	AGGAGGGCAGTTCCGACTTACA
Galnt13	CTGGCGAGAATAAAGGAAGACAG	TTGAGGGACAGGATACCATCGG
Gapdh	TGTGTCCGTCGTGGATCTGA	TTGCTGTTGAAGTCGCAGGAG
Gm5148	GTTTGTCATGCCACTTCTGGTCC	AGATGGCTGGTGAGACTCCACA
Gng4	TTGTGGTGTAGACGCTGGAGCA	CCAGTGTTGTCAGACTTCTTGCC
Grp	AGGCACGGTCCTGGCTAAGATG	CCGTCTCTGTCAGCCGCATAC
Hprt	AGTGTTGGATACAGGCCAGAC	CGTGATTCAAATCCCTGAAGT
Htr1B	TCACTGACCTGCTCGTGTCCAT	TATCCGACGACAGCCAGAAGTC
Il34	CAGGAGGTTCAGACATTGCTGG	GCAGTTGTCCAGCAAGGCTTTG
Irf4	GAACGAGGAGAAGAGCGTCTTC	GTAGGAGGATCTGGCTTGTCGA
Itga5	ACCTGGACCAAGACGGCTACAA	CTGGGAAGGTTTAGTGCTCAGTC
Jun	CCTTGAAAGCTCAGAACTCGGAG	TGCTGCGTTAGCATGAGTTGGC
Kcna5	AGACTCTGCCTGAGTTCAGGGA	GAATGGGTCAGCCAGTGTCCTA
Klhl14	CTGCTGTTGGTTGGAGGATTGC	CTACCTCTACAACGCAGTGGTG
L Myc	ATGGACTTCGACTCGTATCAGC	AGAAATCCTCTCCGCAGTCATA

Lhfpl5	GCTCCATCATCTGCTTCAGCCT	CCCAACCATCTGGGTAGACTAG
Lin7B	GAGTCGCTTCTGCTCCGCCAT	GCCGTGAAAGCAGCCACTGTG
Ly6H	CCACCGATACCGTTTGTGCCAG	AGTGCCGCTTAACGAAGTCGCA
Mad1	AGATGCCTTCAAACGGAGGAA	GTGAGTTGATCTGCTACTGGTG
Mad3	CCCGATACACTACACTAAGCCT	CAGCTTCTGGATATGCACCCT
Mad4	GGCGACTTCGCAAGGAAGAA	GAGCCTGAGTTTAGCTCGTCT
Mal	ACCTTCCCTGACTTGCTCTTCG	GCCACAAAGCAGAACACAGACAC
Mal2	CCGACATCCTAAGGACGTACTC	CCATCCTTGGAGCAGAGGTAGA
Max	AGCGAGGTTTCAATCTGCGG	ACGTTTTCGTTCCAGTGCATTAT
Max	TGACCTCAAGCGGCAGAATGCT	TTGGCGTTGGTGTAGAGGCTGT
Mdm2	CCGAGTTTCTCTGTGAAGGAGC	GTCTGCTCTCACTCAGCGATGT
Mdm4	GCTCTCGCACAGGATCACACTA	ATGTCGTGAGGTAGGCAGTGTG
Mga	GAGGAGCACCTACCTTCTTTGT	ACGGGCATCTCGATTAGTAACT
Mgp	AGGAACGCAACAAGCCTGCCTA	CTGCCTGAAGTAGCGGTTGTAG
Mlx	GGTCGAGTATGCCTACAGTGA	GACACTACACTCCCCTTGTGG
Mnt	CGGCTGAAGAAGCCAAATCCAG	GATGTTGCGCTTCAGGGTCTCA
Mnt	TCCTGTAGTGACCAATTCCCC	GGCTCCTTAATGCTGAGTCCA
Msh2	GAACAAAGGCGAGTATGAAGAGG	GCGTCTAAGTGAGCCAGCACAT
Msh2	GTCTGTGCCAACAGAGATGCTC	CCGTAACTGTCTGGAAAGCATGG
Mxi1	AACATGGCTACGCCTCATCG	CGGTTCTTTTCCAACTCATTGTG
Myoc	ACCCAGGAGCAAAGAAGGAGAC	CAATCCTCCATGTGCTTTCCTGG
Ndn	CCTCTGGTTTCCAGACATGGTG	ATGGTGTGGAGATTGGTCAGCC
N-Myc	ACCATGCCGGGGATGATCT	AGCATCTCCGTAGCCCAATTC
Nnat	GTGGTGGAGGAAGAGGGTTAAG	CACATTTTGGGGAGGGCTTTCG
Nov	GTCACCAACAGGAATCGCCAGT	GTAGGTGGATGGCTTTCAGGGA
Nptx2	GGCAAGCCAACGAGATTGTGCT	TGCCATCACTGACAAACAGCGG
Nrip2	GATTTCTGCTGGATGTCTCAGGC	GCTGTTCTACCTGAGTTGGTGG
Otof	CGCCAAAGTCAAGACCCTCTTC	CACAGGAAGTCCTTTCGCTGCT
P21	AGGTGGACCTGGAGACTCTCAG	TCCTCTTGGAGAAGATCAGCCG
P27	GGAGCGGCAAGATCGCATCTC	TCCTGACAAGCCACGCAGTAGA
Pcdh1	CTGACGGCAGCATAGGTGAA	AGTTTGAGGGCGCTGCTCTT
Pcdh7	AACGATAACACGCCCACCTT	GCCATTGCGACCAAAGTCAC
Pcdh9	TTTGATCAACAGCCGCTTTC	CTCCGGAGTTTCCACGATAT
Pcdh10all	GAGCAGGAACATGGCACTTT	AGTAAGGGGTCCGTGAGTTG
Pcdh10long	ACAGTGGTCATGGAGACAGT	TCCTCGGTGCAGTTGGAGAA
Pcdh10short	CCAAGTCAATTCTGCTCTGG	GTGCCAGAGGAACTTCATTC
Pcdh11	CGGAGAACTCGGCTATAAAC	CAATGACATCAAGTCCGAAA
Pcdh17	TCAGTGCCAGAGGAGCAAGG	AGATGCGGTGCGGAGTTCTC
Pcdh18	AGCATAGGGGCTAAAATTGA	CAGCACGTCCACTTCAATGT
Pcdh19	TTGACCGAGACCTGCTGTGC	TGGGAGCATTGTCGTTCAGG
Pcdh8	CTCAGTGGCCCAGAGCAAGA	CCGGAGCAGAGAGCTGTTGA
Pdlim7	CCTCCTGCTATGATGTGCGCTA	GCACAGGTGAAGCAATGGACGT
Pdyn	CTGTGTGCAGTGAGGATTCAGG	GAGACCGTCAGGGTGAGAAAAG
Pkp2	CAGGTGCTGAAGCAAACCAGAG	GACACTCTCTGTCAAGGTGAGC

Pnck	GCTGTCTCCTACCTTCATAGCC	GCCTAGCATGTTGCCAGCTTGT
Pou4F1	AGTACCCGTCGCTGCACTCCA	GACACCGCGATGTCCACGGC
Ralyl	TGAGTGAGCGTCATGCAAGAGC	GGGTCTCTTACTTCCAGGCTTTG
Rest	CATCTAACGCGACACATGCGGA	CCCGTTGTGAACCTGTCTTGCA
Rgag4	AGCCTCGCTCACTGGAGTTCAA	GTCATGTGCTGCCTCTGATCCT
Rgs10	CCTGGAGAATCTTCTGGAAGACC	CTGCTTCCTGTCCTCCGTTTTC
Rgs4	CTGAAGTCGGAATACAGCGAGG	CTGGTGCAAGAGTCCAGGTTCA
Rgs9	TACGGCGATCAGTCCAAGGTCA	CTGCATCCAGTACATAGCGGTG
Rnf152	AGATGAGGACCAGCCAGAAGGA	GTGTGGTATGGCAATGACTGCC
Rpl13a	CCTGCTGCTCTCAAGGTTGTT	TGGTTGTCACTGCCTGGTACTT
Rpp25	TCCACGACAACTTGGCTACCAG	CTTAGCGGTGCCTTCTTCAGCA
Ruvbl2	GGATACCAAGCAGATTCTGCGC	GTGATGAGCTGGATGGCATAGC
Scfd2	CTTCCTGGCATTTGAAAGGCTCC	GTTGAGGTCGTCACTGGTTCTC
Scn3A	TCCGAGCCTTATCCCGCTTTGA	GAAGATGAGGCACACCAGTAGC
Sdha	CTTGAATGAGGCTGACTGTG	ATCACATAAGCTGGTCCTGT
Serpina3N	CAACCAGAGACCCTGAGGAAGT	AGGACATCCTCCAGGCTGTAGT
Sfrp5	GAGATGCTGCACTGCCACAAGT	TGCTCCATCTCACACTGGGCAC
Slitrk6	GCTTCACTTGGGAAACAACCGC	TGTGGAGACCAAGGAACATGCC
Smco3	GACGGAGAAGATTGCCATCGTG	TGCCAGTTGTGACACTTGAGCC
Smg1	CTGCTCAGGTTGCTAGTGAAGC	CGCACATACACTTCAGGGTGGT
Srf	CTCACCTACCAGGTGTCGGAAT	CTGCTGACTTGCATGGTGGTAG
St8Sia6	CATGAGCTACGAGGTGGAAAGC	GGATAGTCCACAAAAGGCTGCG
Stk32B	TGTTCAAGGTGGAGCGTGTCCA	GGTAGGTCATGCTCTGGATGTC
Svep1	CCAGATGGTTGGGAATCCTGTG	GGCTGAGTAGAAGCCATTCTCC
Syt16	AGTGACTCTGGTTCTGGAGCCA	GAGACAGTGACTGCGTGGATGA
Tmie	CCAAGGAGACTGTGGTGTTCTG	ATACCGAGCCTCGATCTCCTTC
Ubc	AGCCCAGTGTTACCACCAAG	ACCCAAGAACAAGCACAAGG
Usp1	AGTAGCGTCACACCTGTGGACA	GCTTTCACATTCCAAACACCGAG
Usp12	AACAGCACACCAGACCCAACCT	TGTTCCACGTCAACAGAAAGGTC
Usp15	CAGATGGACACCTACGTTGCTG	TCATCTGGCTCGTCGGTTTCCA
Usp21	TCCTGAATGCCGTGCTACAGTG	AGAGGGCACCAATCACATCTGC
Usp33	AACAGCCAGGAGACCGTCAAAG	GGCTTGCTGATAACCGTGGACT
Usp46	CACAGCATTGCCACGCAGAAGA	AACTCGTGAGCATCCTGCTGCA
Vip	GATGCCGTTTGAAGGAGCAGGT	GAAGTCTGCTGTAATCGCTGGTG
Vstm2L	GACTGGACTGACAAGCAGACGT	AGCTTGTGGGAGATGTTGCTGC
Vwa5B2	CCCTGGAGTTTATGAGGTGGCA	GGAAGTAAGCCTGTCCTCTGCT
Wdfy1	AGAGTGCAGTCACTGTGCTACC	CTGCTCACACTTCTGACAGGAG
Yb1	CAGGAGAGCAAGGTAGACCAGT	TGCTGACCTTGGGTCTCATCTC
Znf477	TTCCACCAGAGGAGCCTGCTAT	TACACAGGAGGCAGCGGTAAGA

V.2.2.4 RNA Sequencing

RNA samples were prepared in the lab from cerebellum of 3 week old mice. After brain dissection, cerebellar tissue was used for RNA extraction using TRIzol/chloroform during homogenation of the fatty brain tissue. Subsequently, the Aurum Total RNA mini Kit (Bio-Rad) was used according to the manufacturer's protocol. RNA dissolved in elution buffer was checked for quality and concentration, frozen to -80°C and shipped to the VIB Nucleomics Core (VIB, Belgium) for RNA sequencing by Next Generation Sequencing (NGS) technology, i.e. massive parallel sequencing using Illumina Truseq procedures on an Illumina Hi-Seq apparatus. Two experiments were run at different time points and an internal control was included in the second experiment to normalize and allow merging of the results with the first experiment. Samples were analyzed for purity and integrity and then used for the experiments. An Illumina system was used for the sequencing. (More info can be found at the company web site http://www.illumina.com/techniques/sequencing.html). To assess the accuracy of a sequencing platform, the Phred quality score (Q score) to accuracy was used. A value of Q has to be equal or higher than 20 to be used for next generation sequencing experiments. The relative A, C, G, T content has been checked and Diversity in Read Quality calculated by the ShortRead 1.20.0 package from Bioconductor (http://www.bioconductor.org). Different controls were performed by the Nucleomics Core to confirm the efficiency of the samples to perform in the sequencing process. In the read pre-processing, the data were transferred in fastq-files and were first processed for technical artefacts: low quality ends (<Q20) were trimmed with FastX 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and the reads that were shorter than 35 bp were excluded. Other reads were excluded after the cut of the adapter which are trimmed only at the end (at least 10 bp overlap and 90% match) with Cutadapt 1.2.1[3]. PolyA-reads (more than 90% of the bases equal A), ambiguous reads (containing N), low quality reads (more than 50% of the bases < Q25) and artifact reads (all but 3 bases in the read equal one base type) were then filtered using FastX 0.0.13 and ShortRead 1.16.3. Extra contaminant was removed using Bowtie 2.1.0.

The preprocessed data from the two experiments were aligned to the reference genome of *Mus musculus* (mus musculus GRCm38.73), for a total of 23 samples. The alignment to the reference genome was performed with Tophat v2.0.8b[4]. An extra filter for quality was added with Samtools 0.1.19 able also to sort the reads from the alignment according to the chromosomes and to index the resulting bam-files.

The annotation of genomic features required different steps: first, the unique transcript-related features associated with a gene were identified and selected from the Cufflinks v2.1.1 toolkit

resulting in gtf-format results[5]. Then the exon chains of transcripts of the same gene were merged with MergeBed from the Bedtools v2.17.0 toolkit[6] for the identification. Gffread from the Cufflinks v2.1.1 toolkit was used again this time to extract the genomic sequence of the identified transcripts: for each transcript the sequence length and percentage of G/C-content were described. The entire identified transcript group was compared with the reference transcripts. All the annotations were received in in gtf-format, fasta-format and tab-delimited text.

Expression levels were defined by counting the reads that overlap specific genes in different steps. Htseq-count 0.5.4p3 was used to count the overlapping reads in the alignment and the results were merged with the reference gene annotation previously made. Genes with less than 1 counts-per-million were excluded. Due to the relevance of the GC content in a sample, samples were normalized and corrected for this value using EDASeq package from Bioconductor[7]. Samples were normalized for library size and RNA composition as well in the same way. FPKM values are the rate of fragments per base multiplied by a number in proportion to the number of fragments sequenced ant is used to simplify the results. FPKM values were generated by dividing each normalized count by the total number of counts and then the scaled counts were divided for each gene by the gene length (in kbp).

V.2.3 Quantitative RT-PCR

V.2.3.1 Pcdhs in brain and lung tissue

Pcdh expression levels were tested in mice brain and in lung tissue. We performed qRT-PCR experiments for Pcdh1, Pcdh7, Pcdh9, Pcdh10, Pcdh11X, Pcdh18 and Pcdh19. Pcdh1 was highly expressed in lungs, together with Pcdh18. All Pcdhs were expressed in brain, as expected (**Figure 77**).

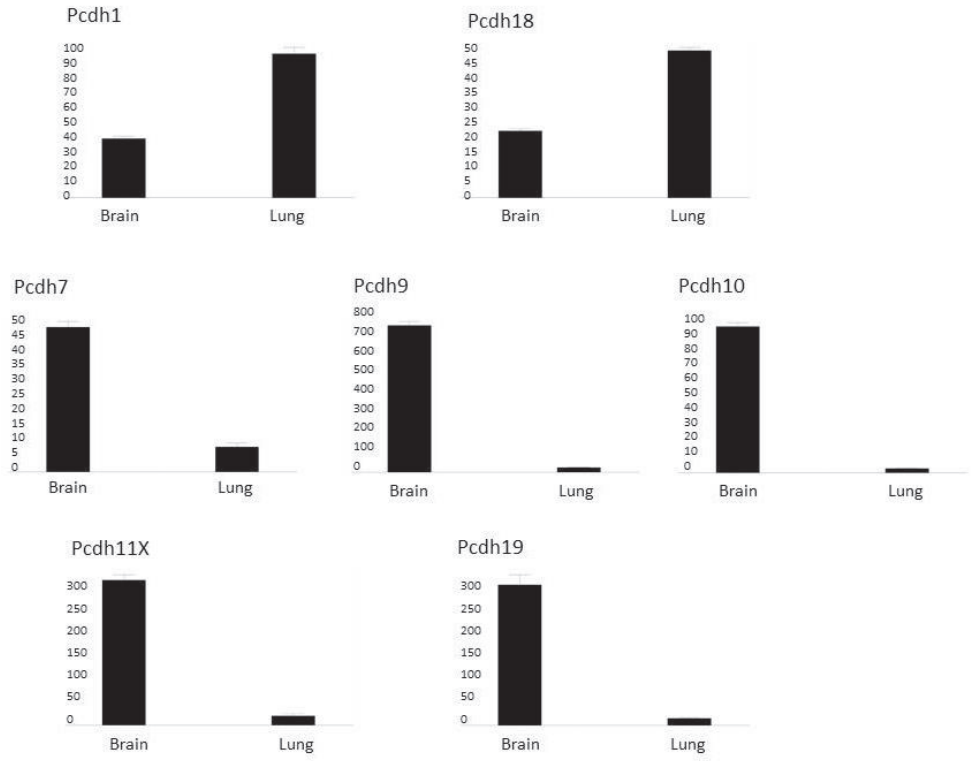


Figure 77 Pcdhs are expressed in mouse brain
Each graph shows the expression of Pcdhs in relative quantities (Y axis). X axis shows the samples: brain on the right and lung on the left.

V.2.3.2 Expression levels of the candidate δPcdh interactors

Myc Target genes

QRT-PCR experiments have been performed and repeated in different settings and samples to assess changes in RNA level of the candidate δPcdh interactors, in lysates of total brain tissue and/or cerebellum, from WT and Pcdh10 KO mice. Preliminary tests were conducted to optimize the experiments and select the most efficient primer pair to amplify each selected gene. Those first experiments, although only preliminary, already showed a tendency towards no or only little variation in expression levels in tissues with or without Pcdh10 (**Figure 78**).

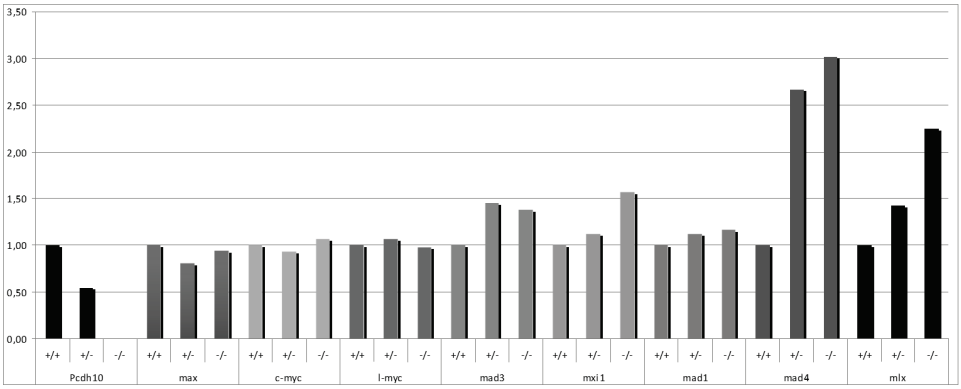


Figure 78 qRT-PCR experiment in mouse brain for Pcdh10 and MAX related candidates
qPCR Primer Pairs were designed for SYBR Green-based real-time qPCR. Total brain was used from 3 week old mice Pcdh10$^{+/+,}$ Pcdh10$^{+/-}$ and Pcdh10$^{-/-}$. We used 1 mouse for each condition. The black bars show Pcdh10 differential expression levels (in correspondence with the Pcdh10 genotypes of the mice analyzed, as indicated by +/+, +/- or -/-), the other bars represent expression levels of different genes in the brains of those mice. A differential expression level in function of the Pcdh10 genotypes was clear only for Mad4 and Mlx.

Since many components of the MAX-MAD pathway were confirmed in a biochemical assay to be δPcdh interactor candidates, expression level experiments were performed for an expanded list, including the MAX-MAD pathway genes and some Myc target genes. In 2008, Kim and colleagues identified over 1400 direct Myc target genes in HeLa cells and human primary fibroblasts, with different roles in cellular pathways and functions[8]. We selected a number of Myc target genes to form a list of candidates on the basis of two studies reporting genes to be involved in prostate and brain cancer[9,10]: some were involved in cell cycle regulation such as *CCNB1, CDC2L1* and *CEB1*, but also in transcription such as *MNT* and *TIP48,* and in DNA repair mechanisms (such as *BRCA1, MSH2, YWHE*)[9]. Finally, also direct MAD3 target genes with involvement in medulloblastoma were included in our selection, such as *BMP3*, *FBL* and *SCFD2*[10].
The experiments were performed on RNA from total brain lysates from 3 week old mice, 3 wild type (+/+) and 3 homozygous (-/-) Pcdh10all mutant mice. Unfortunately, we could not confirm any up- or down-regulation.

The same experiments but with an expanded list were performed on new RNA samples from cerebella lysates of 3 wild type (+/+), 4 homozygous (-/-) mutant Pcdh10all and 4 homozygous (-/-) mutant Pcdh10long mice. Results were analyzed with +QBase, represented via Graph-pad and shown in **Figure 79**, **80**, **81**. Only Ruvbl2 turned out to be downregulated, and this only in the Pcdh10long$^{-/-}$ mice samples (**Figure 80**).

Multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary
WT vs. KO long	-4.917	-12.24 to 2.406	No	ns
WT vs. KO All	7.911	0.5882 to 15.23	Yes	*

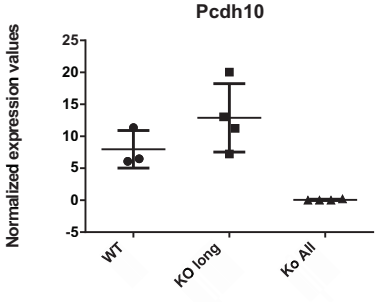


Figure 79 PCDH10 mRNA expression levels in WT, PCDH10all KO mice and PCDH10long KO mice cerebellum
*Ns P value > 0.05, * P value ≤ 0.05, ** P value ≤ 0.01*

Multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary
WT vs. KO long	4.477	-0.6102 to 9.565	No	ns
WT vs. KO all	-2.553	-7.640 to 2.535	No	ns
KO long vs. KO all	-7.030	-11.74 to -2.320	Yes	**

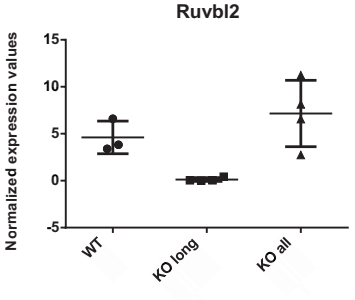


Figure 80 Ruvbl2 mRNA expression is downregulated in PCDH10long KO mouse cerebellum.
*Ns P value > 0.05, * P value ≤ 0.05, ** P value ≤ 0.01*

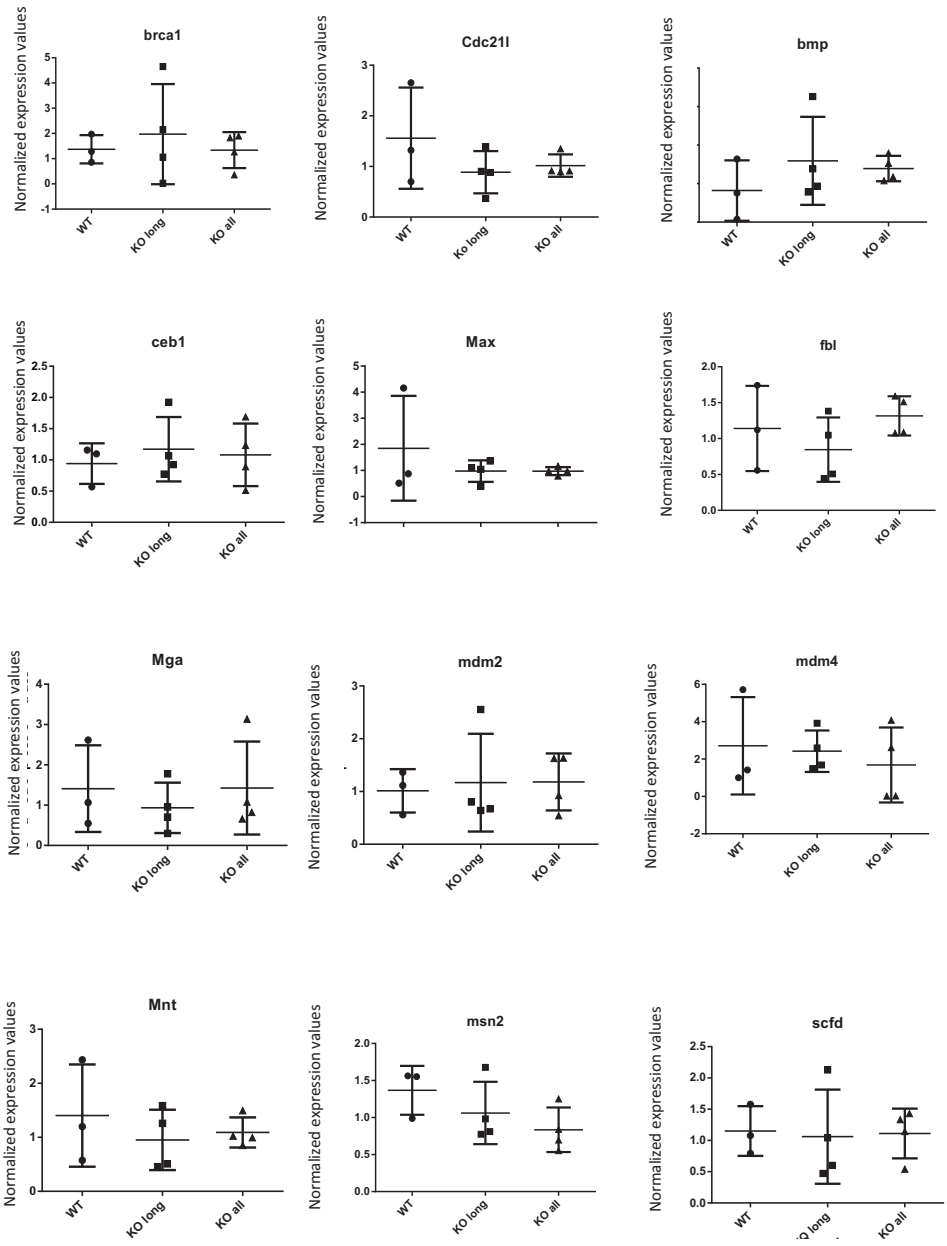


Figure 81 mRNA expression levels of MYC-related genes in Pcdh10all and Pcdh10long KO mouse cerebellum. *QRT-PCR analysis of cerebellum from Pcdh10all and Pcdh10long KO mice. None of the candidates showed differential expression compared with WT mice. X axis= samples, Y Axis= Normalized expression values*

USPs related proteins

Expression levels were investigated in 3 week old mouse cerebella. Ubiquitin-specific proteases (USPs) form a superfamily of deubiquitinating cysteine proteases and are responsible for the cleavage of ubiquitin molecules from proteins and other molecules[11]. Any change in their expression levels in cerebella with or without Pcdh10 could play a role in the homeostasis of the proteins. Since DUB inhibitors are considered as potential anti-cancer agents, we investigated whether Pcdh10 levels are instrumental in the expression levels of USP proteins. QRT-PCR experiments were performed for USP1, USP12, USP15, USP21, USP33 and USP46 in the cerebella from 3 $Pcdh10^{+/+}$all and 3 $Pcdh10^{-/-}$all mice. We could not, however, show any difference in USP expression levels between the KO and the WT mice (not shown).

FHL family members

FHL1, FHL2 and FHL3 belong to the family of the 'Four and a half LIM domain' proteins. The expression levels of the three members of the family did not turn out to be influenced by the Pcdh10 levels in mouse brain tissue (**Figure 82**).

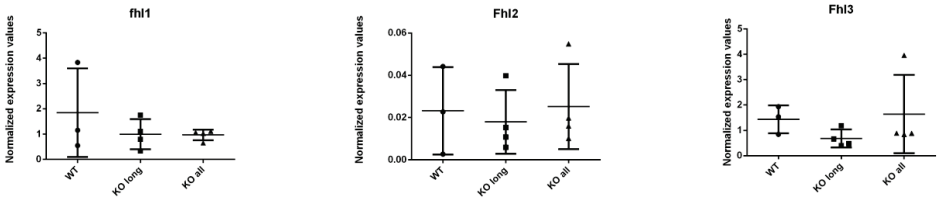


Figure 82 Expression levels of Fhl family members are not influenced by Pcdh10

Other candidates

The expression levels of a list of candidates have been checked, again without showing any difference between Pcdh10 KO and WT mice (examples shown in **Figure 83**).

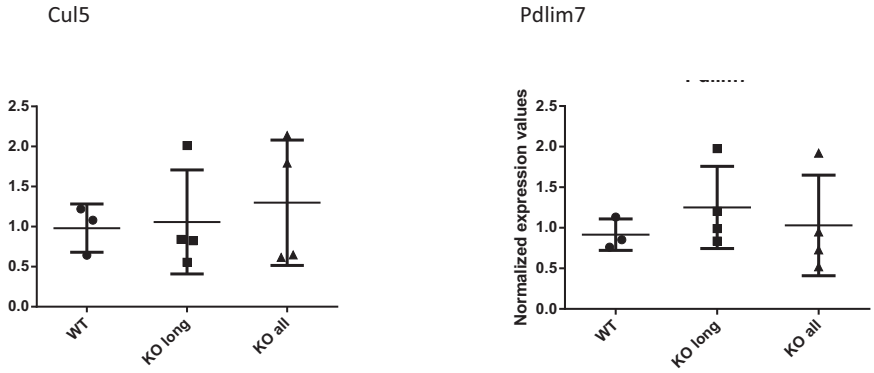


Figure 83 mRNA expression level of Cul5 and Pdlim7 in Pcdh10all and Pcdh10long KO mice cerebellum

V.2.3.3 Conclusions

To investigate whether δPcdhs play a regulatory role in determining the expression level of the respective candidate interaction partners and related genes, qRT-PCR experiments were performed. Pcdh1, Pcdh7, Pcdh9, Pcdh10, Pcdh11X, Pcdh18 and Pcdh19 expression levels were first analyzed in brain and lungs of WT mice, and high expression was found in brain as expected. Then, the gene expression levels of the different candidate interactors were investigated in total brain or cerebellum of WT, Pcdh10all KO and Pcdh10long KO mice. Deubiquitinating enzymes (DUBs) are proteases responsible for the cleavage of ubiquitin and are thus modulators of protein degradation. Considering their implication in the regulation of critical pathways such as DNA damage repair, gene transcription, cell cycle progression and apoptosis, the correlation of defective DUB expression with neurological disorders and cancer is not surprising. post-translational protein modifications(including phosphorylation and ubiquitination) are involved in cancer development, for example in breast cancer where such modifications are considered to be mediators of cancer signaling pathways[12–20]. USPs are a class of DUB proteins, which have been shown to play a role in regulated signaling pathways implicated in cancer: for example, USP15 has been found to be amplified in human breast cancer, ovarian and glioblastoma tumors and is an important regulator of cell cycle progression[13]. Therefore, we analyzed whether changes in the expression level of different members of the USP group can occur in mouse cerebellum as a function of Pcdh expression. Analysis of USP1, USP12, USP15, USP21, USP33, USP46 (all putative interaction partners of Pcdhs) were performed but showed no differences in expression in Pcdh10all KO or Pcdh10long KO mouse cerebellum in

comparison with WT cerebellum. The gene expression of another candidate, CUL5, which is related to the ubiquitin system for protein degradation, was analyzed in the same experiment and did not show a significant difference between KO and WT cerebella either.

PDLIM7 gene expression was studied as well: PDLIM7 is a member of the PDZ-LIM family and contains an N-terminal PDZ domain, three C-terminal LIM domains and a canonical PY (PPXY) motif. It is considered to be an adaptor to stabilize membrane-associated signaling complexes. The PDZ-LIM family is involved in many biological functions such as cardiac and skeletal muscle development and maintenance, neuronal signaling, and tumor cell growth. PDLIM7 associates with actin filaments in fibroblasts via its PDZ domain. It has been shown to negatively regulate p53 through inhibition of MDM2 self-ubiquitination and has been shown to promote tumor cell survival in mice[21]. Evidence has been shown in different studies for the role of PDILM7 in cancer; for instance, its expression level is related to the survival rate of breast cancer patients[22]. Furthermore, PDLIM7 interacts with the receptor tyrosine kinase RET and is responsible for the activation of mitogenic signaling of Ret[23]. Since Ret binds to clustered Pcdhs in differentiated neuroblastoma cells and is required for stabilization and differentiation-induced phosphorylation of Pcdhs11[24], PDLIM7 roles were investigated at different levels, including gene expression. QRT-PCR experiments, performed as previously explained, showed that Pcdh10 expression does not influence PDLM7 gene expression.

Members of FHL family (FHL1, FHL2 and FHL3) are expressed in a cell- and tissue-specific manner and play flexible roles in different cellular systems: for instance, FHL proteins are involved in the regulation of cellular proliferation, differentiation, and adhesion/migration in different tissues and cell lines. Up- or down-regulation of FHL protein levels has been shown in a number of different kinds of cancer, indicating a role of the family in cancer-related pathways. For instance, FHL2, the better-studied member of the family, has been shown to induce EMT in colon cancer as it negatively regulates E-cadherin expression and hence also the formation of E-cadherin–β-catenin complexes[25]. The clinical relevance of the members of the family is apparent from the role that the members play in cancer but mechanisms that regulate those functions are still unclear. I investigated the gene expression levels of the FHL family with and without ablation of Pcdh10 and concluded that in cerebellum Pcdh10 does not play a role in regulation of the RNA levels of members of FHL2 family. Since most evidence for FHL-PCDH interactions has been found in relation to Pcdh11X/Y it would be interesting to investigate whether the silencing or the overexpression of Pcdh11X/Y can influence FHL family members at the RNA level. Studies in vivo have not been performed because of the difficulty in obtaining Pcdh11X KO mice (unpublished data by Dr. Uta Fuchs).

Expression of MAX-MAD pathway genes and some Myc target genes was studied as previously explained. The relevance of this pathway in cancer is nowadays obvious but all the related mechanisms behind this enormously branched influence in cancer are still not entirely clarified: MYC activation has been implicated in the pathogenesis of more than half types of human tumors; it is regulated at different levels but mostly via alterations in signaling pathways that induce or repress its transcription. On the other hand MYC inactivation has been shown to induce rapid apoptosis and also this function requires the binding to MAX. For instance, in lymphoma and in liver adenocarcinoma, MYC inactivation induces proliferative arrest, differentiation/senescence, and widespread apoptosis (reviewed in[26]).

Gene expression experiments were performed to study the role of Pcdh10 in the regulation of RNA levels of members of the MAX-MAD pathways and Myc target genes. Preliminary tests showed an increase of the expression levels of MAD4 and MLX in total brain of Pcdh10all KO mice, but this could not be confirmed in cerebellum, where only RUVBL2 turned out to be influenced by the expression of Pcdh10. Reptin/RUVBL2 and the related protein Pontin/RUVBL1 belong to the AAA+ (ATPases Associated with various cellular Activities) family. They can heterodimerize and show ATPase and helicase activities. Myc binds specific sites in genomic DNA and induces histone acetylation and it has been shown to associates with TIP60 and recruits it to chromatin *in vivo* with four other components of the TIP60 histone acetyltransferase (HAT) complex: TRRAP, p400, RUVBL1 and RUVBL2[27–30]. RUVBL1/2 are involved in many important cellular processes, such as chromatin remodeling, transcription and mitosis, however, the mechanism mediating their function is still unclear. The both interact with β-catenin and c-Myc. RuvBL1 is an essential mediator of c-Myc oncogenic transformation and the interaction between RUVBL1 and cMYC have been shown to influence important MYC functions. So far no functional role has been established for RUVBL2 in MYC functions. C-MYC recruits both RuvBL1 and RuvBL2 for transcriptional control of target genes critical for cell transformation.

Upon Wnt signaling stimulation, the interaction of β-catenin with RUVBL2 in the nucleus induces the transcription of a metastasis suppressor gene (KAL1). If RUVBL2 is sumoylated the transcription of KAL1 is inhibited leading to increased metastatic potential of the cell line. RT-PCR analysis performed on KAI1 mRNA using metastatic LNCap cells, showed that re-expression of RUVBL2 restores KAI1 expression while downregulation persists in case of increased sumoylated Reptin/RUVBL2. RUVBL2 interact also with ATF2, an important inhibitor of the cell cycle which induces apoptosis in response

to DNA damage. Both Reptin/RUVBL2 and Pontin/RUVBL1 expression has been found to be increased in a large number of cancer types showing their role in tumor biology.

In our studies, performed in brain of WT, Pcdh10all KO and PCDH10long KO mice, RUVBL2 mRNA is downregulated in cerebrum of Pcdh10long KO mice indicating a role for the long isoform of Pcdh10 in RUVBL2 expression. Pcdh10 is expected to function as a tumor suppressor gene and RUVBL2 expression is found to be increased in a large number of cancers. A correlation between the known functions of the two candidates and the downregulation of RUVBL2 in absence of longPcdh10 isoform is not clear. The delicate balance between opposite functions played by MYC might be relevant to understand the putative correlation between RUVBL2 and PCDHs. It would be interesting to investigate if the interactions between RUVBL2 and β-catenin or c-MYC are influenced by PCDHs and if the mechanism correlates with the novel interaction partners from the MAX-MAD pathway. Studies of migration and proliferation might be use for this.

V.2.4 RNA Sequencing

V.2.4.1 Report and raw data

RNA samples were prepared in the lab and shipped to the Nucleomics Core facility in Leuven. Each sample was checked twice for quality and purity. The platform used was HiSeq (Illumina). The experiment included 17 samples in total (**Table 22**). The reference organism used was *Mus musculus*. From the preliminary tests, samples showed the required high quality. The mapping efficiency after the filtering of the mapped reads was higher than 97%.

Table 22 Samples used for RNA Sequencing
Samples were obtained from cerebellum of 3 weeks old Pcdh10KO and WT mice, specifically: 7 WT (3 males + 4 females), 5 Pcdh10longKO (3 males + 2 females) and5 Pcdh10allKO (3 males + 2 females). Sex genotyping of mice was performed as previously described (MMRRC's genotyping protocol including Sry primers, www.mmrrc.org/strains/10318).

Exp 1809	pcdh10_all_WTzaa1	WT	MALE
	pcdh10_long_KO_M8	LONG KO	FEMALE
	pcdh10_all_KO_a1	ALL KO	MALE
	Pcdh10_long_KO_I2.5	LONG KO	MALE
	Pcdh10_all_KO_zaa4	ALL KO	MALE
	Pcdh10_long_I13_WT	WT	FEMALE
	Pcdh10_long_KO_M2	LONG KO	FEMALE
	Pcdh10_all_WT_a6	WT	FEMALE
	Pcdh10_long_WT_I1_3	WT	FEMALE
	Pcdh10_WT_zaa3	WT	FEMALE
	Pcdh10_all_KO_zaa6	ALL KO	MALE
Exp 1928	UFVR_07ZAA1C	WT (CONTROL)	MALE
	UFVR_08ZAA5	WT	MALE
	UFVR_09I.2.6	LONG KO	MALE
	UFVR_10ZAA10	ALL KO	FEMALE
	UFVR_I1 3.9	ALL KO	FEMALE
	UFVR_12 I2.2	LONG KO	MALE

To analyze expression levels a Relative log Expression (RLE) plot was generated using a normalized log2-scale count for each gene. The ratio between the expression level of a gene and the median

expression of the same gene in every sample of the experiment was calculated. Only few genes are expected to have different expression levels, which is why, as expected, all the results are similar and in a range close to 0 after normalization. Computational analysis investigating the Spearman correlation and using normalized counts as expression levels produced a heat map of the correlation matrix (**Figure 84**).

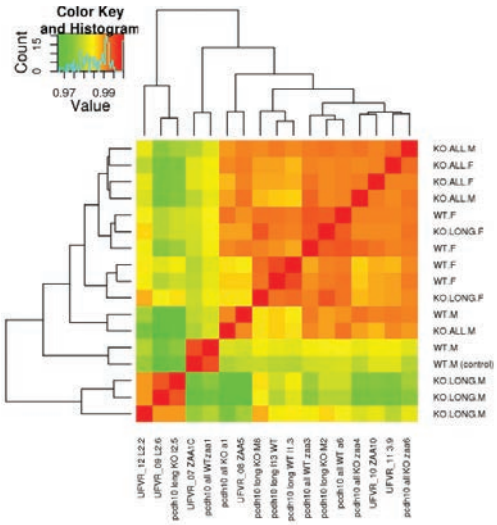


Figure 84 A heat map of the correlation matrix
The columns and rows of the correlation matrix are reordered such that similar samples form clusters. The presence of clusters with a biological interpretation is an indication that normalization successfully removed technical artefacts.

In the heat map, each row and column is associated with a sample and samples with a similar gene expression profile are located close to each other. Correlation between expression profiles is represented where rows and columns cross. This is shown on a color scale: red means high correlation, green means less correlation. The two repetitions of the WT.M sample (analyzed in both exp1809 and exp1928) cluster together: this indicates that the batch effect between the two experiments is probably negligible. This is further confirmed by the fact that samples from both experiments mix in the clustering as expected. Furthermore the heat map shows that samples from the same gender tend to cluster together. In the same way, samples from the same genotype cluster together. Anyway, clustering efforts are generally not perfect, but only show a tendency that samples from the same genotype or the same gender are more similar to each other.

In our investigation mouse gender was considered, and analysis of the results was conducted in gender-related groups. When we compared data from samples of different genders, we only found differences in gender-related genes as being highly significant.

The results of the RNA sequencing were presented as raw data (before filtering and normalization) and as FPKM (fragments per kb of exon per million reads mapped), computed from filtered and normalized counts. Different analysis approaches were used to understand the meaning of the data.

V.2.4.2 Pcdh10 exons in KO mice: validation of genotype in RNA Seq experiment

The number of reads of each *Pcdh10* exon in the RNA Seq experiment on Pcdh10allKO and Pcdh10longKO cerebella was analyzed as an extra control for the KO condition. Exon 1 (**Figure 85**) which contains the start codon was floxed in order to conditionally knockout all isoforms in Pcdh10allKO mice (see above). On the other hand, to preserve expression of the short isoform, exons 2 and 3 were floxed in Pcdh10longKO mice: this causes a frameshift in exons 4 and 5 resulting in expression of the short isoform only. As expected the number of reads for exon 1 demonstrated their absence in Pcdh10allKO mice and their conservation in Pcdh10longKO.

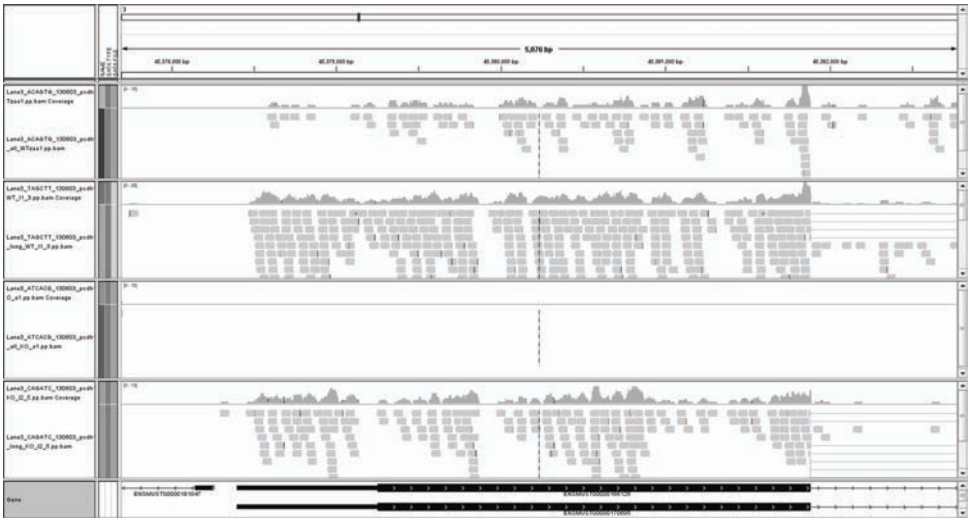


Figure 85 Analysis of reads for exon 1 in Pcdh10all and Pcdh10long KO mouse cerebellum.
Analysis of differential usage of exon 1 in PCDH10, by bam-files in IGV and visually comparing the coverage of the exons. As expected, Pcdh10allKO mice do not have reads for exon 1 (line 3) while the same exon is conserved in Pcdh10long KO mice

V.2.4.3 Software analyses

Data obtained from the sequencing were analyzed and studied for correlations, with Ingenuity® Pathway Analysis (IPA®) (http://www.ingenuity.com/) and MetaCore™ system (http://thomsonreuters.com/en/products-services/pharma-life-sciences/pharmaceutical-research/metacore.html). Transcripts in cerebella from Pcdh10all KO mice didn't show particularly significant differences to those of WT mice. Surprisingly, cerebellar transcripts of Pcdh10long KO mice showed many differences upon comparison to those of WT or Pcdh10all KO mice.

Ingenuity® Pathway Analysis

To start with, the eight first-ranked for differential expression transcripts (cut-off of data: FDR.tgw =0.05) in common for Pcdh10long KO compared withPcdh10all KO or WT were analyzed for molecular or functional relationships: RGAG4, POU4F1, SMCO3, MYOC, RGS10, RGS4, GM5148 and MGP. The complete list of transcripts that show a change in expression can be found in **Table 23**.

Table 23 Complete list of the 27 transcripts found differentially expressed in Pcdh10long KO compared withPcdh10all KO or WT (Ingenuity® Systems, www.ingenuity.com)

Gene.Name	Description
5330417C22Rik	RIKEN cDNA 5330417C22 gene
C130046K22Rik	RIKEN cDNA C130046K22 gene
Cartpt	CART prepropeptide
Dkk3	dickkopf homolog 3 (Xenopus laevis)
Etv1	ets variant gene 1
Exosc9	exosome component 9
Galnt13	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 13
Gm11637	predicted gene 11637
Gm5148	predicted gene 5148
Gpr88	G-protein coupled receptor 88
Hoxb5	homeobox
Htr3a	5-hydroxytryptamine (serotonin) receptor 3A
Kcnd2	potassium voltage-gated channel, Shal-related family, member 2
Klhl14	kelch-like 14
Mgp	matrix Gla protein
Myoc	myocilin
Ndn	necdin
Pou4f1	POU domain, class 4, transcription factor 1

Ralyl	RALY RNA binding protein-like
Rgag4	retrotransposon gag domain containing 4
Rgs10	regulator of G-protein signaling 10
Rgs4	regulator of G-protein signaling 4
Rgs9	regulator of G-protein signaling 9
Rnf152	ring finger protein 152
Smco3	single-pass membrane protein with coiled-coil domains 3
Svep1	sushi, von Willebrand factor type A, EGF and pentraxin domain containing 1
Zcwpw2	zinc finger, CW type with PWWP domain 2

From the 27 candidates of the complete list we selected the 8 more convincing in terms of reads and consistency. IPA did not show any direct interaction between the proteins encoded by the genes of this shortlist. An extended analysis was then performed to find connections including intermediate elements. Results are shown in **Figure 86**: Myc, Ubc and Bcl6 appear to be linkage molecules of the Top list of the RNA Seq results. Pcdh1, 8 and 9 are known to be related to these groups of molecules. Nothing is known about the other members of the Pcdh family, including Pcdh10, which relates them to the network

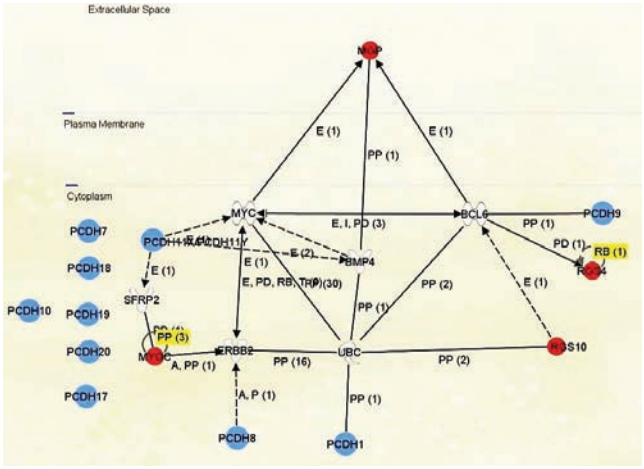


Figure 86 Shortest pathway analysis of the eight first ranked genes differently expressed in the RNA Seq analysis

Only one node was added by the program to analysis to investigate putative correlations between the proposed molecules (red). The pathway is shown considering the subcellular localization. (Ingenuity® Systems, www.ingenuity.com). Only the molecules which show an interaction are shown.

Next, we investigated the role of those 8 molecules and whether they are related to known specific functionIn **Figures 87, 88**, and **89**, their connections to carcinoma, cell movement, and differentiation are shown.

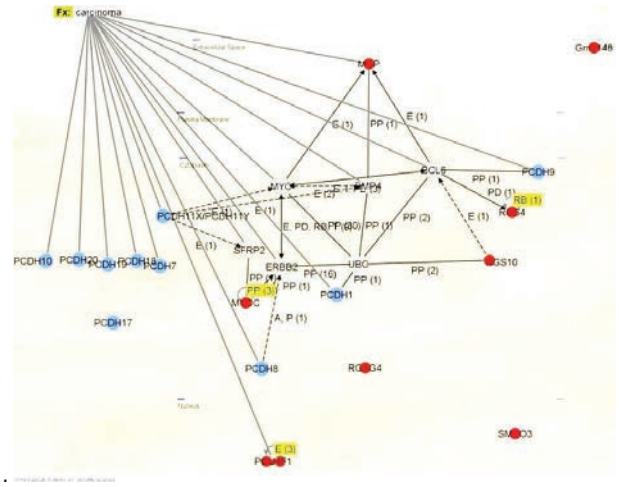


Figure 87 Correlation of δPcdh family members and up/down regulated molecules in Pcdh10long KO mouse cerebellum with carcinoma pathways.

δPcdh are proposed to be proto-oncogenes and tumor suppressor genes. The identification of a carcinoma related pathway might help to elucidate the mechanisms involved in tumor development and progression. (Ingenuity® Systems, www.ingenuity.com). In red: the molecules from the top 8 list; in blue: Pcdhs.

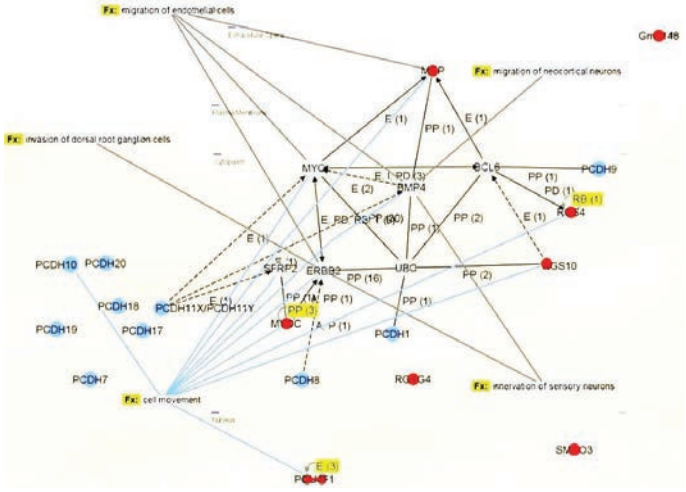


Figure 88 Cell movement related pathways and correlation with δPcdhs and up/down regulated molecules in Pcdh10long KO mouse cerebellum (Ingenuity® Systems, www.ingenuity.com)

In red: the molecules from the top 8 list; in blue: Pcdhs.

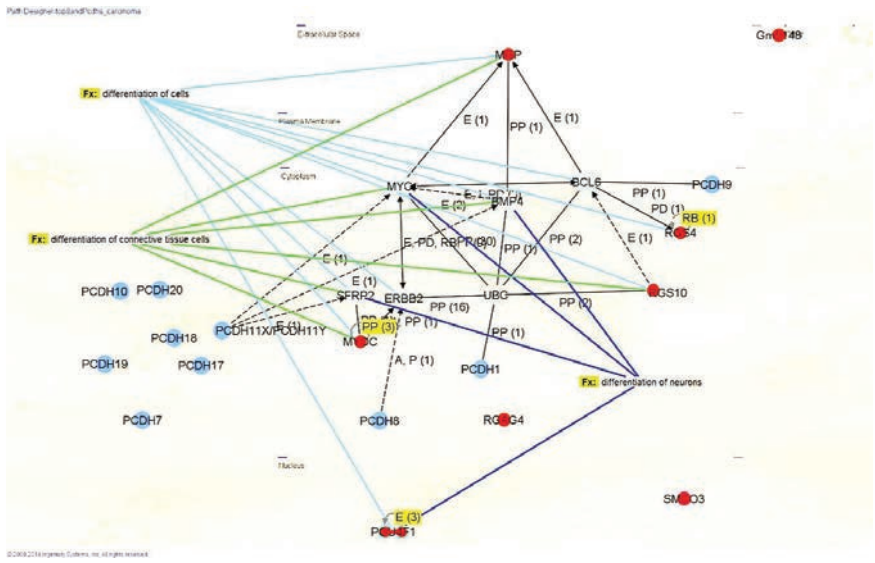


Figure 89 δPcdhs and "top 8 list" molecules and cell differentiation related pathways (Ingenuity® Systems, www.ingenuity.com)

Because of our focus in cancer, we analyzed the carcinoma related connections in detail. PCDH10 has been proposed to be a tumor suppressor gene for nasopharyngeal, esophageal, colorectal, breast, cervical, prostate and other carcinomas, showing frequent methylation of the *PCDH10* gene[31–34]. Furthermore, PCDH10 is involved in inhibition of proliferation, apoptosis induction, and repression of invasion[35]. The IPA analysis found correlations with carcinoma mostly for MGP (Matrix gla protein). Published real-time RT-PCR experiments relate this protein to different epithelial ovarian cancer (EOC) subtypes including mucinous ovarian carcinoma, clear-cell ovarian carcinoma and serous ovarian carcinoma in humans[36], while microarray analysis has shown that an upregulation of MGP mRNA in carcinoma-associated fibroblasts is associated with oral squamous cell carcinoma in human[37]. MGP interacts with BCL6 (B-cell lymphoma 6), a nuclear phosphoprotein related to RGS4 and RGS10, which belongs to our shortlist. Furthermore, BCL6 has been shown to interact with PCDH9 in humans[38]. BCL6 plays a role in carcinoma: for example, downregulation of human BCL6 mRNA in breast tumors is associated with triple-negative breast cancer in females[39], and somatic missense mutations of the *BCL6* gene are associated with adenocarcinoma in human colon[40]. Additionally, PCDH9 is related to carcinoma itself. For example, it has been shown that a somatic mutation of the *PCDH9* gene is associated with adenocarcinoma in human esophagus[41] and the Catalogue Of Somatic Mutations In Cancer (COSMIC) reports an association for *PCDH9* mutations

with adenocarcinoma in human cecum, in human colon and human esophagus. BCL6 interacts with UBC,[42] which is an interaction partner of PCDH1 (our findings).

IPA software was used to compare the RNA levels for cerebella from Pcdh10long KO and WT mice, Pcdh10long KO and WT+Pcdh10all KO mice. In the comparison between Pcdh10long KO and WT, the first 100 higher ranked genes were uploaded and filtered for significant number of reads: only 73 genes were in the filtered list. 41 of these genes are associated with connective tissue development and function, embryonic development and organ development pathways, and 32 with cell morphology, cellular assembly and organization and cellular function and maintenance pathways.

With respect to the main diseases and functions in which the members of the list can be involved, the analysis showed that 20 molecules are involved in neurological diseases, 15 in hereditary disorders and 8 in cancer. 17 molecules are related to cellular development and 9 to cellular growth and proliferation. Many of them are related to physiological system development in embryos and adults. The list of upstream regulators of the pathways includes ARID4B, HOXB3, HTT and REST. REST (RE1-silencing transcription factor) is a transcriptional repressor which binds neuron-restrictive silencer element (NRSE) and represses neuronal gene transcription in non-neuronal cells; it is related to 4 relevant molecules for our study and interacts with UBC (**Figure 90**, **Figure91** and **Table 24**).

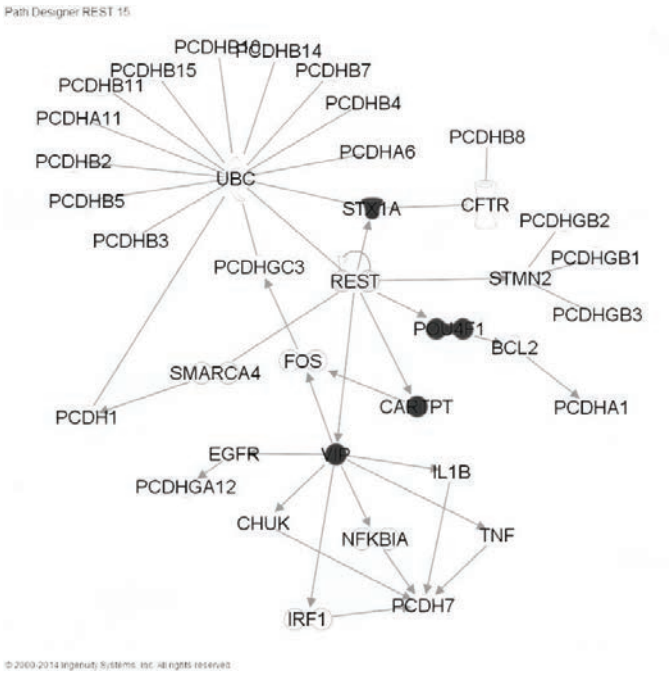


Figure 90 REST and PCDHs (Ingenuity® Systems, www.ingenuity.com)
IPA analysis of the shortest link between REST and PCDHs. Dark molecules= putative candidates from our results POU4F1, VIP, CARTPT and STX1A.

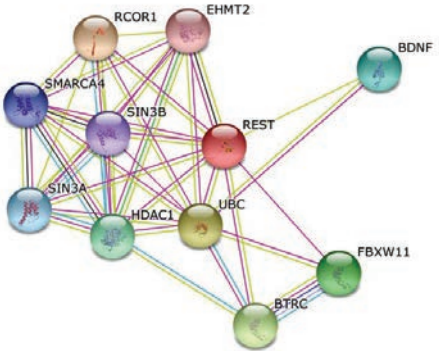


Figure 91 REST interaction partners. Obtained with STRING.
Input: REST; Restricts the expression of neuronal genes by associating with two distinct corepressors, mSin3 and CoREST, which in turn recruit histone deacetylase to the promoters of REST-regulated genes. Mediates repression by recruiting the BHC complex at RE1/NRSE sites which acts by deacetylating and demethylating specific sites on histones, thereby acting as a chromatin modifier (1097 aa). (Homo sapiens)

Table 24 REST predicted functional partners obtained with STRING

RCOR1	REST corepressor 1; Essential component of the BHC complex, a corepressor complex that represses essential component of the BHC complex, a corepressor complex that represses transcription of neuron-specific genes in non-neuronal cells. The BHC complex is recruited at RE1/NRSE sites by REST and acts by deacetylating and demethylating specific sites on histones, thereby acting as a chromatin modifier. (482 aa)
UBC	ubiquitin C (685 aa)
BTRC	beta-transducin repeat containing E3 ubiquitin protein ligase; Substrate recognition component Substrate recognition component of a SCF (SKP1-CUL1-F-box protein) E3 ubiquitin-protein ligase complex which mediates the ubiquitination and subsequent proteasomal degradation of target proteins. Recognizes and binds to phosphorylated target proteins. SCF (BTRC) mediates the ubiquitination of CTNNB1 and participates in Wnt signaling. (605 aa)
FBXW11	F-box and WD repeat domain containing 11; Substrate recognition component of a SCF (SKP1-CUL1-F-box protein) E3 ubiquitin-protein ligase complex which mediates the ubiquitination and subsequent proteasomal degradation of target proteins. Probably recognizes and binds to phosphorylated target proteins. SCF (FBXW11) mediates the ubiquitination of phosphorylated CTNNB1 and participates in Wnt signaling. (542 aa)
HDAC1	histone deacetylase 1; Responsible for the deacetylation of lysine residues on the N-terminal part of the core histones (H2A, H2B, H3 and H4). Histone deacetylation gives a tag for epigenetic repression and plays an important role in transcriptional regulation, cell cycle progression and developmental events. (482 aa)
BDNF	brain-derived neurotrophic factor (329 aa)
SIN3A	SIN3 transcription regulator homolog A (yeast); Acts as a transcriptional repressor. Corepressor for REST. Interacts with MXI1 to repress MYC responsive genes and antagonize MYC oncogenic activities. Also interacts with MXD1-MAX heterodimers to repress transcription by tethering SIN3A to DNA (1273 aa)
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4 (1679 aa)
SIN3B	SIN3 transcription regulator homolog B (yeast); Acts as a transcriptional repressor. Interacts with MXI1 to repress MYC responsive genes and antagonize MYC oncogenic activities. Interacts with MAD-MAX heterodimers by binding to MAD. The heterodimer then represses transcription by tethering SIN3B to DNA. (1162 aa)
EHMT2	euchromatic histone-lysine N-methyltransferase 2 (1210 aa)

The same procedure and parameters were applied to compare Pcdh10long KO and Pcdh10all KO mice. Notably, the results were similar to the ones for the comparison with the WT so a final analysis was performed with Pcdh10long KO vs. WT + Pcdh10all KO.

For this final analysis the first 300 molecules were uploaded (cutoff of logRatio=1 and FDR=0.05). Five main associated network functions were identified by the analysis:

- Behavior, Cell-To-Cell Signaling and Interaction, Molecular Transport; 50 molecules
- Behavior, Endocrine System Disorders, Organismal Injury and Abnormalities; 30 molecules
- Lipid Metabolism, Molecular Transport, Small Molecule Biochemistry; 29 molecules
- Neurological Disease, Organismal Injury and Abnormalities, Dermatological Diseases and Conditions; 25 molecules
- Hereditary Disorder, Neurological Disease, Psychological Disorders; 25 molecules

Concerning diseases and disorders, not surprisingly, the analysis showed high impact in Psychological Disorders (41 molecules) and Neurological Diseases (48 molecules). Besides this the most represented function is Cell-To-Cell Signaling and Interaction and many molecules turned out to be involved in Nervous System Development and Function. The top up- and down- regulated molecules are listed and described in **Table 25** and **Table 26.**

Table 25 Up- regulated and down-regulated molecule top list obtained with IPA analysis system and ranked according to the expression value
Pcdh10long KO vs. WT + Pcdh10all KO mice. The top 300 molecules were analyzed (cutoff of logRatio=1 and FDR=0.05).

Molecules	Exp. Value	Exp. Chart
COMP	↑1.952	
ZCWPW2	↑1.574	
TGM5	↑1.105	
CERKL	↑1.079	
FZD7	↑0.686	
GABRA6	↑0.683	
CNR1	↑0.498	
HOXB5	↓-5.905	
CHRNA5	↓-3.959	
HTR3A	↓-3.537	
CARTPT	↓-3.443	
QRFPR	↓-3.395	
VIP	↓-3.227	
POU4F1	↓-3.075	
GPR88	↓-2.939	
PRPH	↓-2.721	
CRHBP	↓-2.525	

Table 26 Description of up- and down-regulated molecules, obtained via STRING.

Molecule	Description
TGM5	transglutaminase 5; Catalyzes the cross-linking of proteins and the conjugation of polyamines to proteins. Contributes to the formation of the cornified cell envelope of keratinocytes (720 aa)
COMP	cartilage oligomeric matrix protein; May play a role in the structural integrity of cartilage via its interaction with other extracellular matrix proteins such as the collagens and fibronectin. Can mediate the interaction of chondrocytes with the cartilage extracellular matrix through interaction with cell surface integrin receptors. Could play a role in the pathogenesis of osteoarthritis. Potent suppressor of apoptosis in both primary chondrocytes and transformed cells. Suppresses apoptosis by blocking the activation of caspase-3 and by inducing the IAP family of survival proteins. (757 aa)
HOXB5	homeobox B5; Sequence-specific transcription factor which is part of a developmental regulatory system that provides cells with specific positional identities on the anterior-posterior axis (269 aa)
PRPH	peripherin; Class-III neuronal intermediate filament protein (470 aa)
CRHBP	corticotropin releasing hormone binding protein; Binds CRF and inactivates it. May prevent inappropriate pituitary-adrenal stimulation in pregnancy (322 aa)
GABRA6	gamma-aminobutyric acid (GABA) A receptor, alpha 6; GABA, the major inhibitory neurotransmitter in the vertebrate brain, mediates neuronal inhibition by binding to the GABA/benzodiazepine receptor and opening an integral chloride channel (453 aa)
FZD7	frizzled family receptor 7; Receptor for Wnt proteins. Most of frizzled receptors are coupled to the beta-catenin canonical signaling pathway, which leads to the activation of disheveled proteins, inhibition of GSK- 3 kinase, nuclear accumulation of beta-catenin and activation of Wnt target genes. A second signaling pathway involving PKC and calcium fluxes has been seen for some family members, but it is not yet clear if it represents a distinct pathway or if it can be integrated in the canonical pathway, as PKC seems to be required for Wnt-mediated inactivation of GSK-3 kinase. (574 aa)
CARTPT	CART prepropeptide; Satiety factor closely associated with the actions of leptin and neuropeptide y; this anorectic peptide inhibits both normal and starvation-induced feeding and completely blocks the feeding response induced by neuropeptide Y and regulated by leptin in the hypothalamus. It promotes neuronal development and survival in vitro (116 aa)
CHRNA5	cholinergic receptor, nicotinic, alpha 5 (neuronal); After binding acetylcholine, the AChR responds by an extensive change in conformation that affects all subunits and leads to opening of an ion-conducting channel across the plasma membrane (468 aa)
GPR88	G protein-coupled receptor 88; Orphan receptor (384 aa)
CERKL	ceramide kinase-like (558 aa)
HTR3A	5-hydroxytryptamine (serotonin) receptor 3A, ionotropic; This is one of the several different receptors for 5- hydroxytryptamine (serotonin), a biogenic hormone that functions as a neurotransmitter, a hormone, and a mitogen. This receptor is a ligand-gated ion channel, which when activated causes fast, depolarizing responses in neurons. It is a cation-specific, but otherwise relatively nonselective, ion channel (516 aa)
VIP	vasoactive intestinal peptide; VIP causes vasodilation, lowers arterial blood pressure, stimulates myocardial contractility, increases glycogenolysis and relaxes the smooth muscle of trachea, stomach and gall bladder (170 aa)
CNR1	cannabinoid receptor 1 (brain) (472 aa)
POU4F1	POU class 4 homeobox 1; Probable transcription factor which may play a role in the regulation of specific gene expression within a subset of neuronal lineages. May play a role in determining or maintaining the identities of a small subset of visual system neurons (419 aa)
ZCWPW2	zinc finger, CW type with PWWP domain 2 (356 aa)
QRFPR	pyroglutamylated RFamide peptide receptor; Receptor for the orexigenic neuropeptide QRFP. The activity of this receptor is mediated by G proteins that modulate adenylate cyclase activity and intracellular calcium levels (431 aa)

In the list of upstream regulators, together with POU4F1 and HTT, which were already included in the list of down- regulated molecules, as expected, we notice the presence of REST. CACNA1H, CARTPT, EPHA10, POU4F1, STX1A, TAC1 and VIP are the target molecules in our dataset for REST and CARTPT, GAL, GPR64, HOXB5, NECAB2, POU4F1, PRPH, TAC1 are the targets of POU4F1. When analyzed for function and disease, as expected and previously anticipated from the Pcdh10long KO and WT comparison, the main roles played by the bigger fraction of differentially expressed molecules are related to neurological disorders. A few examples are listed in **Table 27**.

Table 27 Diseases and functional annotations related to differentially expressed molecules (Ingenuity® Systems, www.ingenuity.com)

Category	Diseases or Functions Annotation	Molecules
Behavior	behavior	ADRA1B,BCL11B,CARTPT,CELF6,CHRNA5,CRHBP,DRD2,DRD5,EPHB6,GAL,GFRA1,GPR26,GRP,LRRC7,NPTX2,OPRK1,PAK6,PCDH8,PDE11A,PDYN,POU4F1,RGS9,RXRG,SLC30A3,SLITRK6,SSTR2,STX1A,TAC1,TRHR,VGF,VIP
Neurological Disease	Movement Disorders	ADRA1B,BASP1,BCL11B,BID,CACNA1H,CARTPT,CHRNA5,DKK3,DRD2,DRD5,GABRA3,GABRG3,GPR83,GPR88,HTR3A,KCNA5,MAL2,MYH7,OPRK1,PDYN,PTPN5,RGS4,RGS9,RXRG,SCN3A,SCN3B,SCN5A,SERPINA3,TAC1
Molecular Transport	transport of molecule	BCL11B,CACNA1H,CARTPT,CDH23,CNGB1,CPNE6,DRD2,F2R,FABP5,FXYD7,GABRA3,GAL,GFRA1,GRIP2,GRP,KCNA5,LIN7B,NNAT,PDYN,RSPO1,SCN3B,SCN5A,SLC30A3,SSTR2,STX1A,TAC1,VIP
Neurological Disease	dyskinesia	ADRA1B,BASP1,BCL11B,CHRNA5,DKK3,DRD2,DRD5,GABRA3,GABRG3,GPR83,GPR88,HTR3A,KCNA5,MAL2,MYH7,OPRK1,PTPN5,RGS4,RGS9,RXRG,SCN3A,SCN3B,SCN5A,SERPINA3,TAC1
Psychological Disorders	Huntington's Disease	ADRA1B,BASP1,BCL11B,DKK3,DRD2,DRD5,GABRA3,GABRG3,GPR83,GPR88,HTR3A,KCNA5,MAL2,MYH7,OPRK1,PTPN5,RGS4,RGS9,RXRG,SCN3A,SCN3B,SCN5A,SERPINA3,TAC1
Neurological Disease	Huntington's Disease	ADRA1B,BASP1,BCL11B,DKK3,DRD2,DRD5,GABRA3,GABRG3,GPR83,GPR88,HTR3A,KCNA5,MAL2,MYH7,OPRK1,PTPN5,RGS4,RGS9,RXRG,SCN3A,SCN3B,SCN5A,SERPINA3,TAC1
Hereditary Disorder	Huntington's Disease	ADRA1B,BASP1,BCL11B,DKK3,DRD2,DRD5,GABRA3,GABRG3,GPR83,GPR88,HTR3A,KCNA5,MAL2,MYH7,OPRK1,PTPN5,RGS4,RGS9,RXRG,SCN3A,SCN3B,SCN5A,SERPINA3,TAC1
Skeletal and Muscular Disorders	Huntington's Disease	ADRA1B,BASP1,BCL11B,DKK3,DRD2,DRD5,GABRA3,GABRG3,GPR83,GPR88,HTR3A,KCNA5,MAL2,MYH7,OPRK1,PTPN5,RGS4,RGS9,RXRG,SCN3A,SCN3B,SCN5A,SERPINA3,TAC1
Psychological Disorders	Mood Disorders	ADRA1B,ANKRD6,CACNA1H,CARTPT,CHRNA5,CRHBP,DRD2,DRD5,GABRA3,GABRG3,HTR3A,OPRK1,PDE11A,PDYN,RGS4,SCN3A,SCN3B,SCN5A,SLC30A3,STX1A,TAC1,VGF
Cell-To-Cell Signaling and Interaction	neurotransmission	CARTPT,CHRNA5,CNGB1,CNIH3,CPNE6,CRHBP,DRD2,DRD5,F2R,GABRA3,GAL,HTR3A,OPRK1,PCDH8,RGS4,SCN5A,SSTR2,STX1A,SYT5
Nervous System Development and Function	neurotransmission	CARTPT,CHRNA5,CNGB1,CNIH3,CPNE6,CRHBP,DRD2,DRD5,F2R,GABRA3,GAL,HTR3A,OPRK1,PCDH8,RGS4,SCN5A,SSTR2,STX1A,SYT5

Psychological Disorders	Schizophrenia	ADRA1B,CACNA1H,CHRNA5,CRHBP,DRD2,DRD5,GABRA3,GABRG3,GFRA1,GRIP2,LIN7B,OPRK1,PCDH8,RGS4,RXRG,SCN5A,STX1A,TAC1,VGF
Neurological Disease	Schizophrenia	ADRA1B,CACNA1H,CHRNA5,CRHBP,DRD2,DRD5,GABRA3,GABRG3,GFRA1,GRIP2,LIN7B,OPRK1,PCDH8,RGS4,RXRG,SCN5A,STX1A,TAC1,VGF
Hereditary Disorder	Schizophrenia	ADRA1B,CACNA1H,CHRNA5,CRHBP,DRD2,DRD5,GABRA3,GABRG3,GFRA1,GRIP2,LIN7B,OPRK1,PCDH8,RGS4,RXRG,SCN5A,STX1A,TAC1,VGF
Cardiovascular Disease	Heart Disease	ADRA1B,CACNA1H,CHRNA5,DRD2,DRD5,F2R,GABRA3,GABRG3,KCNA5,MGP,MYH7,OPRK1,PDE11A,RGS4,SCN3B,SCN5A,TAC1,TIMP2,VIP
Cellular Function and Maintenance	organization of cytoskeleton	ADRA1B,BASP1,BCL11B,CACNA1H,DKK3,DRD2,F2R,GAL,GFRA1,GRP,IGSF9,LRRC7,PAK6,POU4F1,PRPH,SLITRK6,TAC1,VGF,VIP
Cellular Assembly and Organization	organization of cytoskeleton	ADRA1B,BASP1,BCL11B,CACNA1H,DKK3,DRD2,F2R,GAL,GFRA1,GRP,IGSF9,LRRC7,PAK6,POU4F1,PRPH,SLITRK6,TAC1,VGF,VIP
Neurological Disease	seizure disorder	ADRA1B,ANKRD6,BID,CACNA1H,CHRNA5,DRD2,GABRA3,GABRG3,GAL,GFRA1,NPTX2,PDYN,RGS4,SCN3A,SCN3B,SCN5A,SSTR2,TAC1
Molecular Transport	concentration of lipid	ADRA1B,BCL11B,BID,CARTPT,DKK3,DRD2,EPHB6,F2R,FABP5,GAL,GRP,OPRK1,RGS4,RXRG,SSTR2,TAC1,VGF,VIP
Small Molecule Biochemistry	concentration of lipid	ADRA1B,BCL11B,BID,CARTPT,DKK3,DRD2,EPHB6,F2R,FABP5,GAL,GRP,OPRK1,RGS4,RXRG,SSTR2,TAC1,VGF,VIP
Lipid Metabolism	concentration of lipid	ADRA1B,BCL11B,BID,CARTPT,DKK3,DRD2,EPHB6,F2R,FABP5,GAL,GRP,OPRK1,RGS4,RXRG,SSTR2,TAC1,VGF,VIP
Nervous System Development and Function	morphology of nervous system	BID,CDH23,CNGB1,CRTAC1,DRD2,GAL,GFRA1,HOXB5,KCNA5,LRRC7,NRARP,OTOF,PAK6,POU4F1,RGS11,RXRG,SLITRK6,TAC1
Cellular Function and Maintenance	microtubule dynamics	BASP1,BCL11B,CACNA1H,DKK3,DRD2,F2R,GAL,GFRA1,GRP,IGSF9,LRRC7,PAK6,POU4F1,PRPH,SLITRK6,TAC1,VGF,VIP
Cellular Assembly and Organization	microtubule dynamics	BASP1,BCL11B,CACNA1H,DKK3,DRD2,F2R,GAL,GFRA1,GRP,IGSF9,LRRC7,PAK6,POU4F1,PRPH,SLITRK6,TAC1,VGF,VIP
Metabolic Disease	glucose metabolism disorder	ADRA1B,CARTPT,CNGB1,CPNE4,DRD2,DRD5,FABP5,GABRA3,GABRG3,GAL,GRP,HTR3A,PDE11A,PDYN,SCN5A,SERPINA3,SSTR2,VIP
Neurological Disease	seizures	ANKRD6,BID,CACNA1H,CHRNA5,DRD2,GABRA3,GABRG3,GAL,GFRA1,NPTX2,PDYN,RGS4,SCN3A,SCN3B,SCN5A,SSTR2,TAC1
Nutritional Disease	obesity	ADRA1B,CACNA1H,CARTPT,DRD2,FABP5,GABRA3,GABRG3,GAL,HTR3A,PDE11A,RXRG,SCN3A,SCN3B,SCN5A,SSTR2,TIMP2,VGF
Psychological Disorders	bipolar disorder	ADRA1B,ANKRD6,CACNA1H,CRHBP,DRD2,DRD5,GABRA3,GABRG3,HTR3A,OPRK1,PDYN,RGS4,SCN3A,SCN3B,SCN5A,VGF
Neurological Disease	bipolar disorder	ADRA1B,ANKRD6,CACNA1H,CRHBP,DRD2,DRD5,GABRA3,GABRG3,HTR3A,OPRK1,PDYN,RGS4,SCN3A,SCN3B,SCN5A,VGF

The IPA system offered a lot of information but definitive conclusions could not be reached. The indications given form the analysis might be useful in the future when new discoveries and information can be integrated. The same data were also analyzed with the Metacore system which calculates p-values for networks generated by the algorithms. P-value calculations are based on hypergeometric distributions and are used to evaluate network's relevance to Gene Ontology (GO) biological process classification.

112 genes from the RNA sequencing experiment were analyzed. The analysis for canonical pathways offered a set of signaling and metabolic maps, which are created by Thomson Reuters scientists by a process based on published peer-reviewed literature (**Figure 92**).

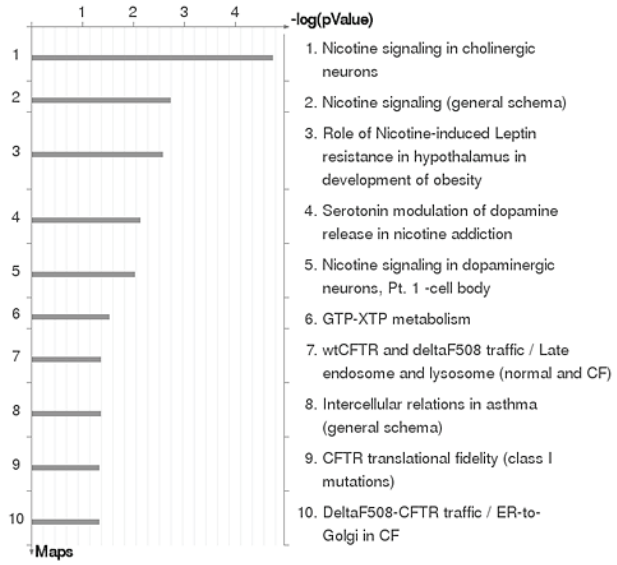


Figure 92 Significance of Pathway Maps.
Sorting is done for the 'Statistically significant Maps'. (MetaCore™ Analysis, Thomson Reuters).

The maps identified by the analysis mainly points to the nicotine signaling including different pathways. Nicotine works by the binding to nAChR receptor: in the brain this is responsible for the adaptation effect that occurs in response to chronic exposure. The main responsible of nicotine addiction is dopamine release. Dopamine binds to the receptor and activates a feedback loop regulation. DARPP-32, dopamine and cyclic AMP-regulated phospho-protein, is one of the key

players in the dopamine signaling: when phosphorylated in a specific threonine residues, it inhibits PPP1CA causing the inhibition of dopamine secretion. Recently several evidence (*in vitro* studies on cell cultures and *in vivo* studies on rodents) indicate a role of nicotine in cancer development[43,44]. Data from the analysis are visualized in **Figure 93** where different nicotine related signaling pathways are shown, including the nicotine signaling in cholinergic neurons which is indicated to be the most significant map from the analysis. Details of this pathway are shown in **Figure 94**.

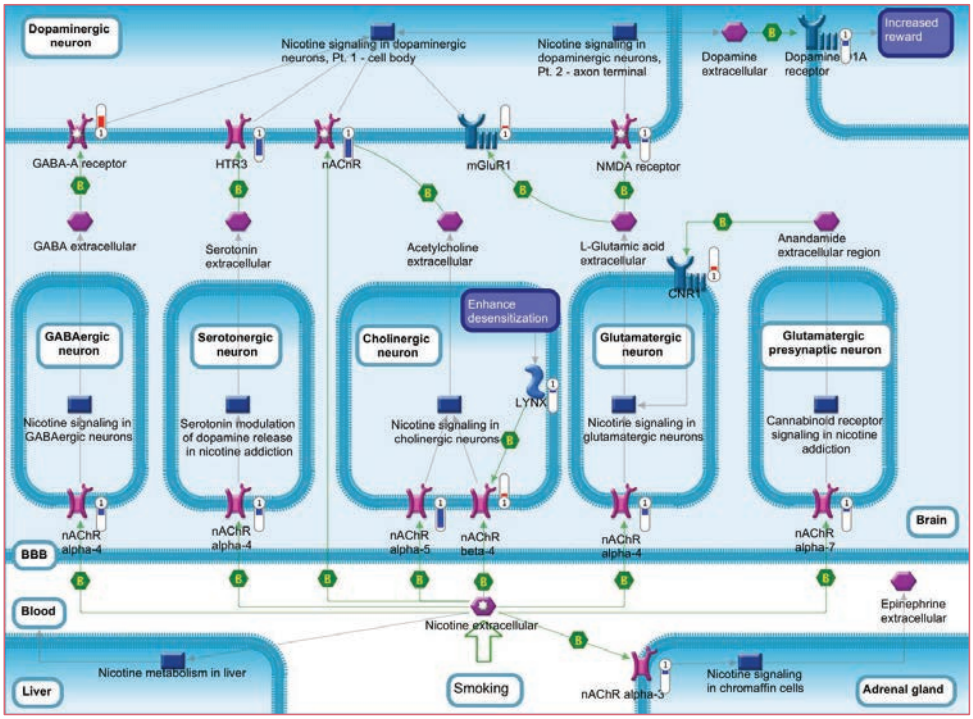


Figure 93 Nicotine signaling. General scheme
Experimental data from all files is linked to and visualized on the maps as thermometer-like figures. Up-ward thermometers have red color and indicate upregulated signals and down-ward (blue) ones indicate downregulated expression levels of the genes. MetaCore™ Analysis, Thomson Reuters

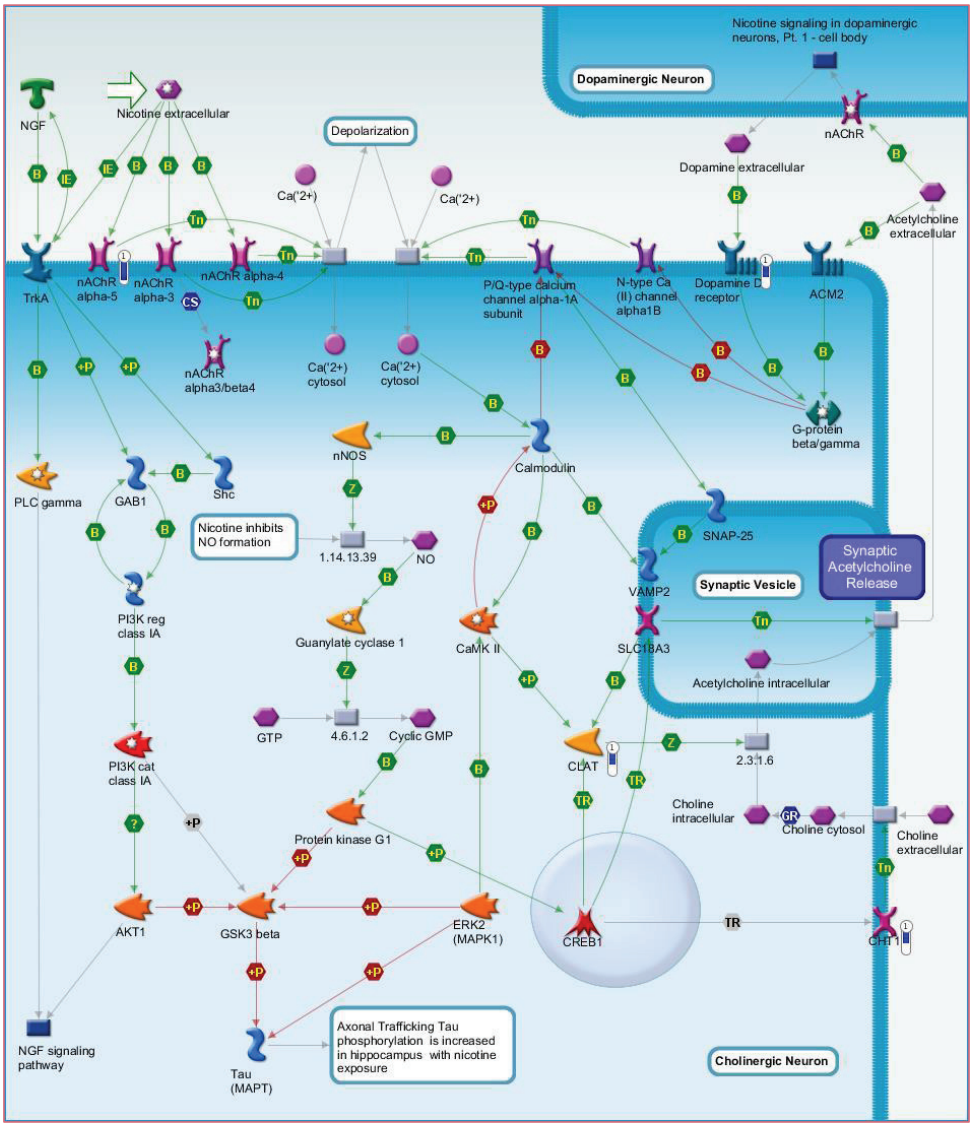


Figure 94 The top scored map (map with the lowest p-value) based on the enrichment distribution sorted by 'Statistically significant Maps' set.

Experimental data from all files is linked to and visualized on the maps as thermometer-like figures. Up-ward thermometers have red color and indicate upregulated signals and down-ward (blue) ones indicate downregulated expression levels of the genes. MetaCore™ Analysis, Thomson Reuters.

MetaCore use the Gene Ontology (GO) classification to investigate the functions of molecules of interest. The GO project aims to describe and correlate genes across different databases in association with biological processes, cellular components and molecular functions in a species-independent manner (**Figure 95**).

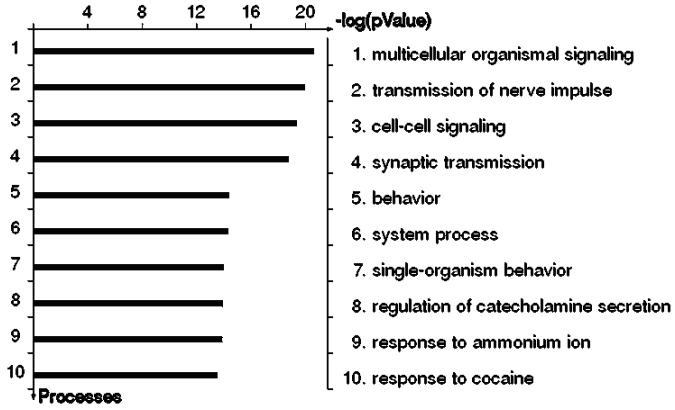


Figure 95 The significance of GO Processes associated with the network
The GO Processes tab is used to see the top 10 functional processes associated with the network. Sorting is done for 'Statistically significant Processes'. MetaCore™ Analysis, Thomson Reuters

We analyzed the data using the Diseases tab function to identify the top 10 diseases associated with the network. The list includes bipolar disorders, affective disorders and psychotic ones, trigeminal neuralgia and nerve diseases and gastrointestinal related diseases (vomiting, constipation and digestive diseases). Three specific networks were proposed to be related to the obtained data. The gene list used as the input data generated a biological network using the Analyze Networks (AN) algorithm: the obtained networks were prioritized based on the number of fragments of canonical pathways on the network (details in **Table 28**).

Table 28 MetaCore™ Analysis, Thomson Reuters

No	Network name	Processes	Size	Target
1	Galpha(i)-specific peptide GPCRs, GRP, Galpha(q)-specific peptide GPCRs, Substance P, Thyroliberin	cell surface receptor signaling pathway (89.1%), signal transduction (97.8%), response to organic cyclic compound (63.0%), response to oxygen-containing compound (69.6%), response to endogenous stimulus (69.6%)	55	5
2	Substance P extracellular region, CART, CREB1, Neurokinin-2 receptor, LRCH4	behavior (61.2%), feeding behavior (40.8%), micturition (26.5%), behavioral response to nicotine (24.5%), regulation of amine transport (34.7%)	50	12
3	BHMT2, Guanine deaminase, DPY19L2, GRP, INDOL1	synaptic transmission (58.0%), transmission of nerve impulse (58.0%), multicellular organismal signaling (58.0%), cell-cell signaling (58.0%), neuropeptide signaling pathway (28.0%)	50	11

In summary, two analysis steps in MetaCore™ (Enrichment Analysis Workflow and Analyze Network) led us to the identification of processes which are differently regulated in the cerebellum of Pcdh10long KO mice. Enrichment analysis and network construction allowed us to evaluate statistically significant genes, which distinguish WT and KO brains. In addition, we examined the implicated disease pathways.

V.2.4.4 QPCR experiments on RNA Seq results

QRT-PCR experiments were performed in cerebrum of WT, Pcdh10All KO and Pcdh10Long KO mice: molecules of interest were selected from the RNA Sequencing experiment and analysis. A list of about 67 genes was compiled. From the list of significant hits obtained by RNA Seq analysis, we selected candidates based on the ranking starting from the highest significance of differential expression. This list of selected transcripts is shown in **Table 29.**

Table 29 List of genes analyzed on qRT-PCR experiments

5330417C22Rik	Ephb6	Myoc	Rpp25
Adcy1	Exosc9	Ndn	Scn3a
Basp1	F2r	Nnat	Sfrp5
Bid	Fndc9	Nov	Slitrk6
Cartpt	Fxyd7	Nptx2	Smco3
Cbln3	Galnt13	Nrip2	Smg1
Cdh23	Gm5148	Otof	St8sia6
Cdon	Gng4	Pdyn	Stk32b
Cerkl	Gpr88	Pnck	Svep1
Cetn4	Il34	Pou4f1	Syt16
Cst3	Kcna5	Ralyl	Tmie
Cul5	Klhl14	Rest	Vip
Cux1	Lhfpl5	Rgag4	Vstm2l
Dact2	Lin7b	Rgs10	Vwa5b2
Dkk3	Ly6h	Rgs4	Wdfy1
Dos	Mal2	Rgs9	Yb1
Eed	Mgp	Rnf152	

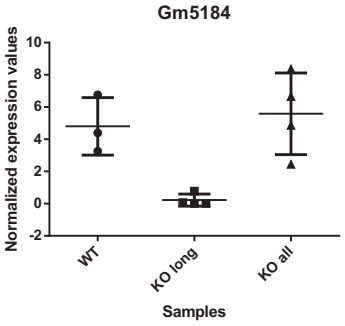
In these experiments, as expected we could not identify statistically significant differences in the expression of the genes (except for Pcdh10) in total KO mice when compared with WT, while the knock out for the long isoform of Pcdh10 showed again some examples of up- and down-regulation when compared with Pcdh10KOall and WT mice, as summarized in **Table 30**. Graphpad software is here used to show the two most significant results as an example to highlight the following

conclusions: in accordance with the RNASeq data, WT brain and Pcdh10all KO brain do not show transcriptional differences, while Pcdh10long KO brain shows statistically significant (*/**) differences when compared with either WT or Pcdh10all KO brain (**Figure 96**). In this experiment we obtained relative quantities of expression for three mice for each condition.

Table 30 Up- and down-regulated molecules in Pcdh10Long KO mice in qRT-PCR experiments

+	Ccnd	*Cyclin D1*
-	Cetn4	*Centrin 4*
-	Dos	*Protein Dos*
-	Epn	*Epyphycan*
-	Exosc9	*Exosome complex 9*
-	Fas	*TNF Receptor Superfamily, Member 6*
-	Gm5184	*Pseudogene*
-	Mgp	*Matrix Gla Protein*
-	Pnck	*Pregnancy Up-Regulated Nonubiquitous CaM Kinase*
-	Rgs4	*Schizophrenia Disorder 9*
-	St8sia6	*ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 6*
-	Stk32b	*Serine/Threonine Kinase 32B*
-	Syt16	*Synaptotagmin XVI*
-	Tmie	*Transmembrane Inner Ear Expressed Protein*
-	Vip	*vasoactive intestinal polypeptide*
-	Wdfy1	*WD Repeat And FYVE Domain Containing 1*

Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary
WT vs. KO long	4.581	0.6428 to 8.519	Yes	*
WT vs. KO all	-0.7818	-4.720 to 3.156	No	ns
KO long vs. KO all	-5.362	-9.008 to -1.717	Yes	**



Tukey's multiple comparisons test	Mean Diff.	95% CI of diff.	Significant?	Summary
WT vs. KO long	3.543	0.6941 to 6.392	Yes	*
WT vs. KO all	-1.233	-4.082 to 1.616	No	ns
KO long vs. KO all	-4.776	-7.414 to -2.139	Yes	**

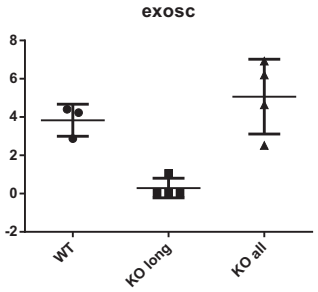


Figure 96 Examples of Downregulation in Pcdh10Long KO mice: Exosc9 and Gm5182 analyzed with Graph Pad.

Gm5184 is a mouse pseudogene otherwise not further described. Exosc9 encodes a component of the RNA exosome which is responsible for processing and degradation of RNA in the nucleus and cytoplasm.

V.2.4.5 Conclusions and discussion

RNA-Seq experiments were performed to identify mRNA expression changes in the cerebellum of Pcdh10all KO mice and Pcdh10long KO mice in comparison with WT mice. Results of those experiments were not easy to interpret.

Cerebellum from WT mice and from Pcdh10all KO mice did not show significant changes in mRNA expression when compared. Surprisingly the comparison of Pcdh10long KO mice with either WT or Pcdh10all KO mice resulted in a long list of up- and down- regulated transcripts. Quite a few of those differences seem to be significant, consistent and reproducible. IPA and Metacore analyses were used to analyze the results and to identify common pathways.

WT vs Pcdh10long KO mice and Pcdh10all KO vs Pcdh10long KO mice analyses were performed, common differently expressed molecules were selected, and these were: Gm5148, Gpr88, Mgp, Myoc, Pou4f1, Rgag4 Rgs10 and Rgs4 mRNAs. Mgp is connected to Rgs4 and Rgs10 through Bcl6 (**Figure 97**). Mgp and Bcl6 correlate with carcinoma in different studies. PCDH1 and PCDH9 interact or correlate with Bcl6. Is not clear how Pcdh10 can play a role in this context.

In conclusion, we have shown that a complicated and difficult to explain change in RNA expression profile occurs in Pcdh10long KO mouse cerebella but not in Pcdh10allKO cerebella. In Pcdh10long KO mice the long isoforms of Pcdh10 are deleted while the short one is conserved. The short isoform does not include the CMs.

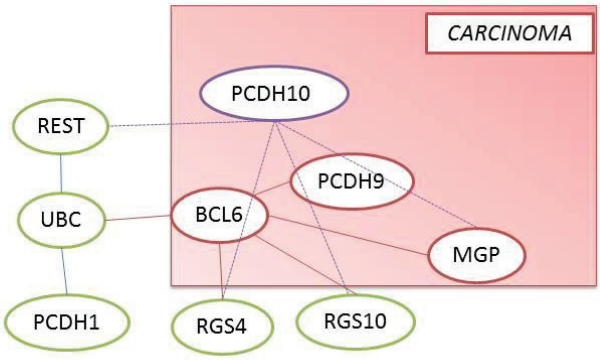


Figure 97 Connections between different expressed molecules indicated from RNA-Seq experiments and PCDHs.

To explain why the total KO mice show the same RNA expression profile as the WT mice, we speculate that other δPcdhs may compensate as they show redundant functions.

QRT-PCR experiments were performed in total brain of Pcdh10allKO mice for δPcdh family members. Statistically significant upregulation was shown for Pcdh7, Pcdh8, Pcdh11, Pcdh17 and Pcdh18 transcripts. When the same experiment was performed in brains of Pcdh10long KO mice, none of these δPcdh showed a change in mRNA expression level (I. Kahr, PhD thesis Ghent University). These results indicate that if Pcdh10 is lost in the very early stages of development, a compensation mechanism occurs based on upregulation of several other members of the protein family. In contrast, it seems that the preserved presence of the short isoform of Pcdh10 does not induce the upregulation of other Pcdhs upon ablation of the long Pcdh10 isoforms; apparently, the short isoform is then upregulated at the RNA level. Western blot analysis in brain lysates of heterozygous and homozygous Pcdh10long KO mice were carried out and as expected these revealed downregulation of long Pcdh10 isoforms in heterozygous mutant and complete absence of the long isoforms in the homozygous KO. Also at the protein level, homozygous Pcdh10long KO mice showed an upregulation of the short isoform. The total Pcdh10 expression level seemed to be similar to the level in the WT brain lysate. Intriguingly, in the RNA Seq analysis the upregulation of other Pcdhs in Pcdh10Long KO mice was not shown. This discrepancy with the QRT-PCR data is puzzling. A possible although rather unlikely explanation might be that on the one hand total brain including cerebellum and on the other hand cerebellum only was used for these two different experiments. It could therefore be interesting to perform extra RNA Seq analyses on total-brain lysate to address this discrepancy.

V.2.5 Nuclear localization

Considering the nature of the identified interaction candidates (mostly transcription factors) and of PCDHs (transmembrane proteins) we considered the possible localization of the interaction: can PCDHs migrate to the nucleus?

Since the localization of a protein is closely related to the specific function of the same protein, many different ways to predict a protein's subcellular localization have been developed. Here we show the results of 8 different computational approaches to study the possibility of PCDH nuclear localization: CELLO2.5,[45,46] Multiloc,[47] PSORTII (http://psort.hgc.jp/form2.html), WoLF PSORT,[48] Hum-mPLoc,[49] NLStradamus,[50] SCLpred[51] and BaCelLO[52]. PCDH sequence was analyzed as full protein, cytoplasmic domain (CD) only or extracellular domain (EC) only, with the aim of identifying the presence of a nuclear localization signal (NLS). An NLS is a sequence which allows the translocation of the protein (or part of it) to the nucleus. The so-called classical nuclear import pathway is the best described

mechanism of transport of protein between cytoplasm and nucleus[53,54]. As shown in **Figure 98**, PCDH11X particularly shows the signs of an NLS in its CDs.

Protein	CELLO2.5	MultiLoc	PSORT II	WoLF PSORT
hSNAI1 (+)	nuclear 4.057	nuclear 0.86 NLS NO	nuclear 78.3%	nucl. 30.0
hCTNND1(+)	nuclear 3.768	nuclear o.98 NLS YES	nuclear 56.6%	nucl. 20.5
hCASP1 (-)	cytoplasmatic 2.762	Cytoplasm o.95 NLS NO	nuclear 39.1%	cyto. 24.5
hPCD10	nuclear 2.451	plasma membr. 0.98 NLS NO	ER 44%	plasma membr. 24.5
hPCD10 CD	NA	nuclear o.88 NLS NO	nuclear 74%	extracell. 26
hPCDH10 ED	NA	plasma membr. 0.95 NLS NO	NA	NA
hPCDH11X	nuclear 2.477	plasma membr. O.89 NLS YES	plasma membr. 35%	Plasma membr. 17
hPCDH11X CD	NA	nuclear 0.95 NLS YES	nuclear 61%	nucl. 27
hPCDH11X ED	NA	cytoplasm o.8 NLS NO	NA	NA

Protein	Hum-mPLoc	NLStradamus	SCLpred	BaCelLo
hSNAI1 (+)	nucleus		nucleus(high reliability)	nucleus
hCTNND1(+)	cytopl.and nucleus		nucleus(high reliability)	nucleus
hCASP1 (-)	cytopl.		Cytoplasm (medium reliability)	cytoplasm
hPCD10	plasma membrane	778-781	secretory (high reliability)	secretory
hPCD10 CD	NA	NA	secretory (medium reliability)	nucleus
hPCDH10 ED	NA	NA	NA	secretory
hPCDH11X	plasma membrane	868-879	Cytoplasm (low reliability)	secretory
hPCDH11X CD	NA	NA	nucleus(high reliability)	nucleus
hPCDH11X ED	NA	NA	NA	secretory

Figure 98 PCDHs showing an NLS, required for nuclear translocation.
8 computational approaches were used to investigate the possibility of nuclear localization for PCDH family members.

Since the separated PCDHs' cytoplasmic domain could display nuclear localization we speculated that these TM proteins could be subject to cleavage phenomena. Some studies have shown that Pcdhs are indeed involved in nuclear signaling. For example, *Xenopus* Pcdh7/NF-Pcdh interacts with chromatin, as well as with TAF1/set, but this last interaction appears to occur in the cytoplasm[55]. Clustered PCDHs have been shown to be cleaved by ϒ secretase in order to produce fragments which can transport information to the nucleus, mostly during development[56]. Further experiments need to be done to investigate the cleavage occurrence in δPCDHs and to identify the localization of the interactions with the candidates. For instance, a strong transcription activator such as VP16 could be used, fused to the full length PCDH protein, to show if events of specific cleavage occur indeed in the CD.

V.3 USP Family Experiments

USP proteins and CUL5 are involved in the proteasome mediated degradation of ubiquitinated proteins. The presence of certain candidates in the list of confirmed interactors for PCDHs required further analysis to study the role of the interaction in the process. We performed different tests to study the possible roles of the interactions.

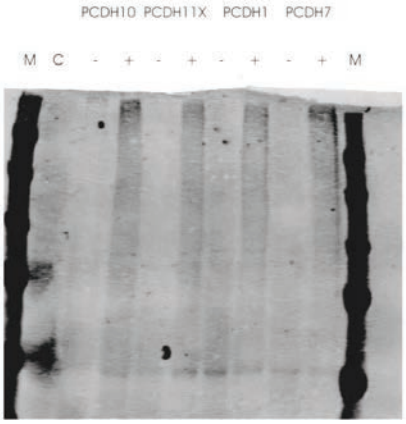
V.3.1 Materials and methods

Ubiquitination assays were performed in HEK293T cells. On day 1 cells were seeded: 2 400 000 HEK293T cells mid-petri dish and transfected after 24h with CaP transfection method as previously described (with 2 μg of total DNA). The day after cells were first washed with PBS and then were incubated over night with MG132 (proteasomal inhibitor, 1/4000) + chloroquine (lysosomal inhibitor, 1/2000) in OPTIMEM medium. The day after cells were washed with PBS, lysed in 250ul 2% SDS gel laemmli buffer (2% SDS, 150 mM NaCl, 10 mM Tris-HCl, pH 8) with 2mM sodium orthovanadate, 50 mM sodium fluoride, NEM and protease inhibitors and sonicated (20 cycles, 10'' on, 5'' off) using the Bioruptor Plus (Diagenode). After boiling for 10', 30 μl are kept aside for total lysate, 2250μl of dilution buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton) were added to the rest. After 1h incubation and 30' of centrifugation, samples were divided concerning the experimental plan and incubated with ab anti-GFP and beads or anti-flag beads. Immunoprecipitation was performed with Dynabeads as previously described. Elution was done in 2X laemmli buffer and samples were boiled 10' before being loaded on SDS-PAGE.

For Lysosome & Proteasome Inhibitor assay 200 000 HEK293T cells were seeded in 12 well plates, transfected, washed, incubate overnight with MG132 OR chloroquine, lysate and sonicated as for ubiquitination assay. After boiling, cell lysates were resolved by SDS-PAGE and transferred to nitrocellulose membranes (Amersham Biosciences). Blotting efficiency was checked by Ponceau S staining (Sigma). Blots were blocked in StartingBlock blocking buffer (Pierce), when using Odyssey infrared imaging (LICOR) or in 5% milk upon enhanced chemiluminescence (ECL) detection. β-actin, HA-tagged, Flag-tagged and E-tagged proteins were revealed using a rabbit or mouse anti-β-actin (1:5000, Sigma), monoclonal rat anti-HA (3F10) (1:5000, Roche), rabbit anti-Flag (1:5000, Sigma) and mouse anti-E-tag (1:10000, Phadia) antibody, respectively, followed by an anti-rabbit or anti-mouse DyLight 800- or DyLight 680-conjugated antibody (1:15000, Pierce) or an anti-rat Alexa Fluor 680-conjugated antibody (1:5000, Molecular Probes), diluted in Odyssey blocking buffer (Li-Cor) + 0.1% Tween-20. Rabbit anti-RNF41 A300-048A (1:20000, Bethyl) or rabbit anti-ASB6 HPA004341 (1:200, Sigma) were revealed by SuperSignal West Pico Chemiluminescent Substrate (Pierce) using peroxidase-conjugated anti-rabbit antibodies (1:5000, Jackson ImmunoResearch).

V.3.2 Results

There is no current evidence of the regulation of PCDHs via ubiquitination so first PCDH ubiquitination was checked. PCDH10, PCDH11x, PCDH7 and PCDH1 were transiently transfected in HEK293T cells. To be able to detect the ubiquitination as a smear in the gel, it is necessary to counteract protein degradation by adding MG132 (proteasomal inhibitor) and chloroquine (lysosomal inhibitor). IP experiments were performed as previously described. Results are shown in **Figure 99**.



0 M			
1(C)	pMet7-sol IL-5R		
2	PCDH10-EYFP	pMet7-sol IL-5R	
3	PCDH10-EYFP	pMet7-sol IL-5R	pcDNA3-HA-ubiquitin
4	PCDH11X-EYFP	pMet7-sol IL-5R	
5	PCDH11X-EYFP	pMet7-sol IL-5R	pcDNA3-HA-ubiquitin
6	PCDH1-EYFP	pMet7-sol IL-5R	
7	PCDH1-EYFP	pMet7-sol IL-5R	pcDNA3-HA-ubiquitin
8	PCDH7-EYFP	pMet7-sol IL-5R	
9	PCDH7-EYFP	pMet7-sol IL-5R	pcDNA3-HA-ubiquitin

Figure 99 PCDHs appear to be ubiquitinated
In presence of proteasomal and lysosomal inhibitors, different PCDHs show that they undergo an ubiquitination process as evidenced by the smear in presence of ubiquitin (+) pMet7-sol Il5R is used as control plasmid. GFP antibody has been used to pull down EYFP tagged PCDHs. Antibody anti HA has been used in the Western Blot.

To investigate the role of different USPs on the ubiquitination level, the same experiment was performed including USP8, USP21 or USP46. Preliminary data show that the tested USPs do not interfere with the ubiquitination of PCDHs (not shown). Extra experiments with a larger panel of USPs are required to confirm this indication.

We performed experiments to analyze the stability of PCDHs and to see if the interaction with USPs could interfere with it. The experiment was performed for PCDH10 and PCD11X with and without USPs at 6 different time points from 0 to 46 hours. In the first 16 hours PCDHs are stable, and then the expression is reduced. We could define a time point of approximately 20 hours where the protein starts to be degraded but this can be due to cellular stress after an extended time post-transfection (**Figure 100**).

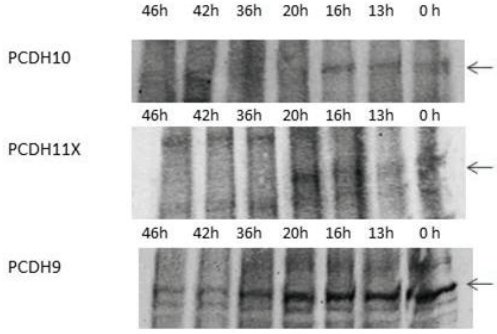


Figure 100 PCDHs show a half-life of about 20 hours
Protein stability experiments for PCDH10, PCDH11X and PCDH9 at 7 different time points. Experiments were performed in transiently transfected HEK293T cells. Incubation with Chloroquine (lysosomal inhibitor, 1/2000).

We performed the same experiment for the first 16 hours to study the role of USPs in the protein degradation. As shown in **Figure 101**, USP8 and USP21 do not interfere with the process. This preliminary experiment will need to be performed again at a bigger scale including different USPs, extra time points (i.e. 20, 36 and 42 hours) and extra positive and loading controls.

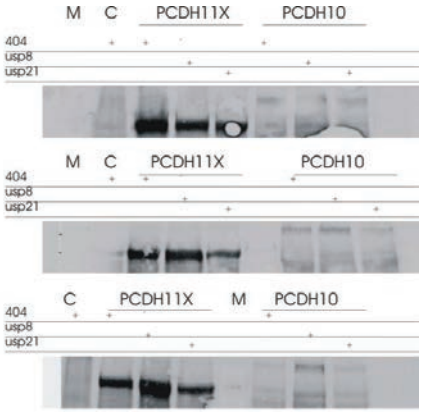


Figure 101 USPs do not interfere with PCDH stability
Experiments were performed in transiently transfected HEK293T cells. Incubation with Chloroquine (lysosomal inhibitor, 1/2000) : first line T0,0h; second line T1,8; third line T2,16h. 404= pMet-sol IL-5R, C= control transfection with 404.

In conclusion, different PCDHs seem to undergo ubiquitination but we could not show any role played by different USP proteins in the process. More experiments are needed to confirm this indication. A more complete panel of USPs should be used, perhaps in the inducible FlpIn Trex HEK 293T PCDHs stable cell lines.

V.4 MAX-MAD pathway

MAX functions as the center of a transcription factor network that includes the MYC family of oncoproteins and the tumor suppressor MAD family. To study the interplay between PCDHs and Myc pathway we used a Cignal Lenti Myc reporter assay. Cignal Lenti Myc reporter kit (SA Biosciences) is designed to monitor the activity of Myc-regulated signal transduction pathways in cultured cells. The reporter kit consists of a Myc-responsive firefly luciferase construct and a construct constitutively expressing Renilla luciferase.

Following the manual guidelines, experiments were performed on inducible HEK293T Flp In TRex PCDH10 EYFP and HEK293T Flp In Trex PCD11X-EYFP cell lines. Induced and not induced cells were transfected with the reporter construct or with controls. The negative control is a mixture of a not inducible reporter construct, which encodes firefly luciferase under the control of a basal promoter element (TATA box) and constitutively expressing Renilla luciferase construct. This was used to measure the reporter background. The positive control was as well taken from the kit and is a constitutively expressing GFP construct together with a constitutively expressing firefly luciferase construct and a constitutively expressing Renilla luciferase construct. This is a control for successful transfection. The Cignal reporter is again a mixture, but this time of an inducible transcription factor responsive construct and the constitutively expressing Renilla luciferase construct. This experiment was rather preliminary as it was performed only in duplicate for each condition, and replicated only once. However, the same result was obtained twice: about no signal of Myc activity for any condition and, consequently, no difference could be demonstrated under the influence of PCDHs.

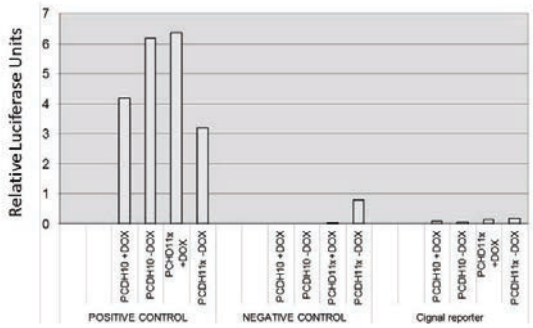


Figure 102 PCDH10 and PCD11X do not interfere with Myc activity in a reporter assay.

Experiments showed that the Myc activity was not regulated by PCDH10 or PCDH11X in a Myc reporter assay (**Figure 102**). The same set up could be used to repeat the experiment after Myc pathway activation, for example by doxorubicin. Furthermore it might be interesting to investigate

the endogenous effects of Pcdh loss on MYC signaling. Pcdh10 KO mice (or cells from there mice) or siRNA against PCDH11 might be use for the purpose. For instance, myc signaling is required for some forms of p53-dependent apoptosis. Specific tests might elucidate the role of this mechanism: if PCDHs activate Myc, apoptosis might increase while If PCDH inactivates MYC, cells may become more resistant to apoptosis. Since Myc is related to many pathways, different studies can still be performed.

V.5 Migration proliferation and adhesion

V.5.1 Introduction

Morphogenetic processes are the foundations of organs and tissue formation during the embryonic development phases and they occur due to the ability of the cells to recognize each other and build stable interactions with other cells and with the extracellular matrix. Generally, in health and disease, these processes, including migration, proliferation and adhesion play important roles for physiologic development and in occurrence and progression of diseases. Due to the presence of EC domains, the first identified function of the δPCDHs subfamily was cell-cell adhesion. On the other hand, considering that the unique intracellular domain of δPCDHs lacks the catenin-interacting motifs which are playing a role in the adhesion function in classical cadherins, this ability has been shown to be weak. Furthermore, some regions of the cytoplasmic domain of δ2PCDHS even show adhesion suppressing functions: for instance, in a fibroblastic cell line, the expression of a mutated PCDH19, which lacks part of the CD results in larger cell aggregates when compared with full-length PCDH19 expressing cells[57]. More recently, a similar behavior has been shown for PCDH17 if expressed in a mutated form in neurons: the lack of the CD induces lateral clumping of axons in embryonic brain[58]. Recent studies have shown that δPCDHs can promote focal cell motility[58–62], while in zebrafish Pcdh18 has been shown to play a role in the reduction of cell migration, maybe due to an enhanced cell adhesion activity[63]. To understand how δPCDHs can play a role in cell motility, different studies have been performed and some hypotheses have been formulated but the exact mechanism is still to be clarified. δ2PCDHs bind different components of the WASP family verprolin-homologous protein (WAVE) complex[57,58,60,61]. This binding occurs via a specific motif, called WIRF and being present in the CD of δ2PCDHs[64]. The WAVE protein is part of a WAVE receptor complex (WRC) formed by Sra1/Cyfip1 (or the orthologue PIR121/Cyfip2), Nap1/Hem2/Kette (or the orthologue Hem1), Abi2 (or the orthologues Abi1 and Abi3), HSPC300/Brick1, and WAVE1/SCAR (or the orthologues WAVE2 and WAVE3)[65]: this complex is responsible for strong actin nucleating activity. Different mechanisms are involved, showing on the one hand the influence of the activation of Arp2/3 (actin-related protein 2/3)[66] or, on the other hand, the role of Abi proteins[67] which allow the binding with lamellipodin (Lpd), a known interaction partner of actin polymerization regulators such as Ena/VASP family proteins. Hayashi and colleagues recently have shown that if the WAVE complex binds Pcdh17, regulation of actin dynamics occurs preferably via the interaction with lamellipodin and Ena/VASP family proteins[58]. In the same study they have shown that in confluent U251 cells the expression of PCDH17, but also of PCDH10, increases the migration of single cells. Those observations brought the authors to the conclusion that Pcdh17 recruits the WAVE-complex,

Lpd and Ena/VASP at cell-cell contact sites and in this way plays a role in the motility of the membrane and thereby promotes cell migration[59].

Interestingly, some of the interaction partners that we identified for different members of the δPCDH family have been shown to play a role in the same mechanisms, such as growth and migration. For example, as it is known and already mentioned, the MAX-MAD pathway is responsible for a delicate balance of the oncogene Myc: while the dimers Myc-Max are involved in cell proliferation and oncogenesis due to their transcription activation function, the Max-Mad dimers act as transcriptional repressor: hence, Mad and Myc are in competition for the binding of Max and the result of the switch from Myc–Max to Mad–Max dimers is a decrease in cell growth and proliferation. In this way the members of the Mad family can inhibit cellular growth and transformation and work as antagonists of the activity of Myc. For example, a role of MAD1 has been shown in suspended or semi solid cultures of the human macrophage cell line U-937: compared to non-expressing MAD1 cells, the sample showed a reduced, almost completely inhibited proliferation, suggesting a role of MAD1 in growth and proliferation of U-937 human monoblast[68].

Concerning another protein family identified as likely interaction partners of δPCDHs, the FHL family, it is known that the proteins of this group play a role in a large number of cellular processes and they take part to the regulation of cell growth, adhesion and migration[69–75]. Furthermore, members of the FHL family have been shown to participate in carcinogenesis and to facilitate the association of proteins with the actin cytoskeleton in the cytoplasm. FHL2 plays a shown role in proliferation but its activity is intriguing since it can have both a positive and a negative effect on cell cycle related processes, depending on the tissues analyzed[76]. Furthermore as already presented in Chapter IV, our experiments indicates actin related protein as potential interaction partners of PCDHs and member of the WRC complex have been already shown to interact with different PCHDs. This class of partners is implicated in cell movement, above all for the role played in the axonal cones and it became of major interest for growth and migration studies.

Given the nature of the identified candidates and based on published data on PCDHs and/or the selected partners, we decided to investigate PCDH dependent effects on cell migration, cell growth and cell adhesion with the final aim of investigating quantitatively the role played by the different novel interactions in various cell properties. We explored functional effects of increasing the level of PCDH10 or PCDH11x in a number of cell-based assays in order to establish a workflow to test the effect of the identified candidate interactors on different PCDHs. The cell-based assay provided a set of tools to address in the future the effect of disturbing the interactions between PCDHs and a selected partner by downregulating this partner by siRNA.

V.5.2 Quantitative analysis of cell properties of cell lines with induced PCDHs expression

V.5.2.1 Methodology

V.5.2.1.1 Cells

HEK293T cell lines expressing EYFP-tagged PCDH10, PCDH11X and PCDH11Y upon doxycycline addition were generated (HEK293T-FlpInTRex PCDH 10-EYPF, HEK293T-FlpInTRex PCDH 11X-EYPF) using the Flp-In™ TRex™ System (Life Technologies) (**Figure 103**). Briefly, the Flp-in™ TRex™ HEK293T host cell line by Invitrogen was used to directly generate Flp-In™ TRex™ expression cell lines. The host cell line has been co-transfected with pcDNA™5/FRT/TO and pOG44 Flp recombinase expression. In this way the gene of interest included in the expression vector is integrated in a Flp recombinase-dependent manner into the genome. pcDNA™5/FRT/TO is a inducible expression vector from Invitrogen and it contains an hybrid human cytomegalovirus (CMV)/TetO2 promoter, a FLP Recombination Target (FRT) site and a Hygromycin resistance gene. These cells were grown in DMEM growth medium supplemented with Glutamax I (Invitrogen) and 10 % fetal bovine serum (FBS, HyClone) at 37°C in 5% CO2. Doxycycline (Sigma) was added to 2µg/ml, 48 h before start of a functional assay. Where indicated, epidermal growth factor (EGF, Invitrogen) was added to 5 nM. HeLa cells stably expressing either PCDH10 tagged to far-red fluorescent protein mKate2, or only mKate2 (control) were grown under the same conditions. These stable cell lines were established by Irene Kahr: HEK293T cells were transiently co-transfected with the pLenti6 (puro)-PCDH10-EYFP construct and packaging plasmids pMD.G and pCMV Delta R8.9 in order to produce the requested lentivirus. The viral supernatant was collected after 48 h, filtered (0.45µm filter by Millipore) and added to HeLa cells. 1 µg/ml Puromycin (Sigma) was added after 48 h for selection. Un-transduced control cultures were used as control: the complete eradication of those supported the 100% transduction of the surviving cells of transduced population.

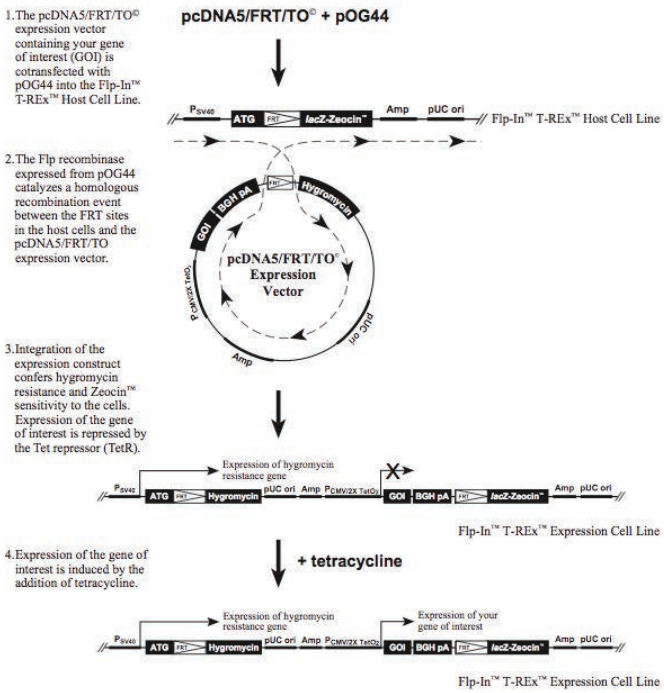


Figure 103 The Flp-In™ System[g]

It is used for generating stable mammalian expression cell lines by Flp recombinase-mediated integration while the T-Rex system[h] is a Tetracycline-Regulated Expression system for mammalian cells: The Flp-in™ T-Rex™ System merges the properties and the protocols of the Flp-In and the T-Rex Systems to obtain a host cell line which contains a single FRT site (introduced by transfection of the pFRT/lacZeo plasmid) and expresses the Tet repressor (introduced by transfection of the pcDNA6/TR plasmid).

---

[g] Briefly, a Flp In target site vector, pFRT/lacZeo, is used to create the host cell line. An SV40 early promoter controls the expression of the fusion gene which is included in the vector, together with a FRT site (as binding and cleavage site for the Flp recombinase) just after its start codon; mammalian cells transfected with the vector can then be first selected for ZeocinTM resistance and then screened for those containing only one integrated FRT site. Once this host cell line is isolated, it can be co-transfected with an expression vector, pcDNATM5/FRT in which we cloned the gene of interest and which carries an FRT site linked to a (defective) hygromycin resistance gene, and with pOG44, a plasmid which constitutively expresses the Flp recombinase. The introduction of the expression vector at the FRT site activates a range of events which ends with the foundation of stable Flp-In™ expression, cell lines, which can be selected for hygromycin resistance, Zeocin™ sensitivity, lack of β-galactosidase activity, and construction of the recombinant protein of interest

[h] The main component of the system is an inducible expression plasmid, in which the gene of interest is cloned in the multiple cloning site and whose expression is regulated by the strong CMV promoter. Two copies of the Tet operating 2 (TetO2) sequence have been inserted in the promoter as a binding site for the Tet repressor. The other component of the system is the pcDNA™6/TR regulatory vector which expresses high levels of the Tet repressor (TetR) gene under the control of the human CMV promoter. Both plasmids are transfected into mammalian cells with a standard method of choice.
In non-induced condition, the transcription of the gene of interest is repressed by the high affinity binding of the repressor to the TetO2 region of the promoter. If tetracycline is added a change of conformation on the homodimers which the repressor form to bind the promoter occurs, with consequent dissociation from it. This allows the initiation of transcription of the factor of interest. (More details can be found on the company website: http://www.lifetechnologies.com).

V.5.2.1.2 Cell exclusion zone migration assay and data analysis

The experimental set-up is shown in **Figure 104**. A 96-well plate (NUNC; Thermo Fisher Scientific) were coated with monomeric rat tail collagen type I (BD Biosciences) by incubating for 1 h at room temperature with 100 µl of 40µg/ml collagen in Ca^{2+} and Mg^{2+} containing Dulbecco's Phosphate-Buffered Saline (D-PBS; Gibco). Fifty-five thousand cells were seeded per well in growth medium in presence or absence of doxycycline and a cell-free area was simultaneously generated in the middle of the well according to the ORIS cell invasion protocol (ORISTM; Plathypus Technologies). We note that the reproducibility of the onset of migration is strongly dependent on the level of confluency at start. After 16 h at 37°C and 5% CO_2 allowing cell adhesion and spreading, the insert and medium was removed and replaced by 200 µl medium containing doxycycline, 1 or 10 % FBS and, where indicated, EGF. For each condition, x-10 replicates (wells) were performed. Subsequently, the cells were allowed to migrate into the cell-free central zone for the indicated time. Phase-contrast time-lapse movies were recorded with a time interval of 20 min using a 10 x UPlanFL objective (N.A. 0.30) on a CellM live cell imaging system with temperature and CO_2-control and an IX81 microscope (Olympus). Image and data analysis were done using CELLMIA, image processing software developed in the group Van Troys/Ampe (UGent, Faculty of Medicine, Dep. Biochemistry) that quantifies the area covered by the cells in time[77], and using the data management and analysis software, CellMissy[78]. The latter derives the median velocity of the cells using linear regression of the area over time curves and statistically compares the different conditions using pair wise non-parametric tests (Wilcoxon) with Benjamin-Hochberg correction for multiple testing.

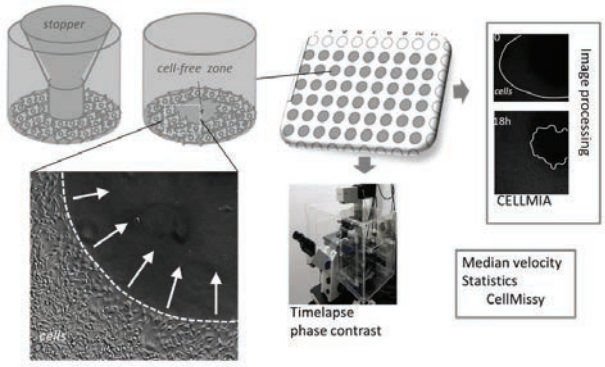


Figure 104 Migration assay work flow.
Migration of cells seeded as a confluent layer is followed in the central cell free zone in time followed by image and data processing. The velocity of the cell layer (median of velocity in replicates) is based on the evolution of the cell covered area in time. 4-6 conditions are tested in parallel with 10 replicates per condition.

V.5.2.1.3 Cell proliferation assay

HEK293T-FlpInTRex PCDH 10-EYFP, HEK293T-FlpInTRex PCDH 11x-EYFP cells were either pre-treated with doxycycline (induction of expression of the fusion protein) or not (control) and seeded at 4500 per well in a 96-wells tissue culture plate (6 replicates per condition per plate, number of plates equals number of time points) that was coated with monomeric collagen type I as described above. After overnight incubation, one plate (T0) was treated according to the procedure described in Cell Proliferation Kit II (Roche). This assay is based on the cleavage of the tetrazolium salt XTT (2, 3-Bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide salt) in the presence of an electron-coupling reagent, producing a soluble formazan salt (**Figure 105**). Mitochondrial enzymes are inactivated shortly after cell death: only living mitochondria can reduce the XTT.

The cells were incubated with the XTT labelling mixture for four hours and formazan production was quantitated using a scanning multi-well spectrophotometer (450 nm). The absorbance directly correlates to the number of viable, metabolically active cells. The measurement was repeated at different time points TX (e.g. T48). The ratio's Tx/T0 (e.g. T48/T0) is indicative of cell growth. The different time points or conditions were statistically compared using a t-test.

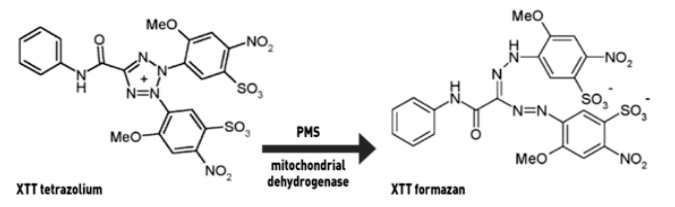


Figure 105 Cleavage of the tetrazolium salt XTT
The tetrazolium salt of XTT (2, 3-Bis-(2-methoxy-4-nitro-5-sulfophenyl]-2H-tetrazolium-5-carboxyanilide salt) is an inner salt, that is cleaved to formazan by the succinate dehydrogenase system of the mitochondrial respiratory chain. The dehydrogenase is still active only in cells with an intact mitochondrial and cell membrane. In the shown example phenazine methosulfate (PMS) is the electron coupling agent which is necessary as an intermediate electron acceptor during the reaction. (Modified from https://www.applichem.com/literatur/applications/nr-12-zellproliferations-assay-xtt/)

V.5.2.1.4 Cell adhesion assay

Cell adhesion kinetics are measured in real time on the iCelligence instrument (ACEA, http://www.aceabio.com) (**Figure 106**), based on impedance when the cells adhere to and spread on the gold coating of the E-Plate L8 (ACEA) in normal growth medium. The wells were coated with collagen (10 µg/ml) as described by the E-Plate manual (ACEA). HEK293T-FlpInTRex PCDH10-EYFP, HEK293T-FlpInTRex PCDH11X-EYFP cells that were pre-treated for 48 h with doxycycline or not

(control), were added at 20000 cells to the wells and measurement (every minute) is immediately started for up to 8 hours (four replicates/condition). Adhesion and spreading is mainly occurring in the first two hours. The assay is sensitive to small differences in the number of cells added at start, and thus to differences in cell number between the populations that are compared. Currently cell counting is done repeatedly in both concentrated and diluted cell suspensions to reduce errors in counting.

The assay system expresses impedance in arbitrary Cell Index (CI) units (Rn-Rb)/4.6; where Rn is the cell-electrode impedance of the well when it contains cells and Rb is the background impedance of the well with the media alone, and the constant 4.6 is related to the voltage used. From average CI versus time (T, in minutes), curves ΔCI/ΔT were determined for the time intervals indicated.

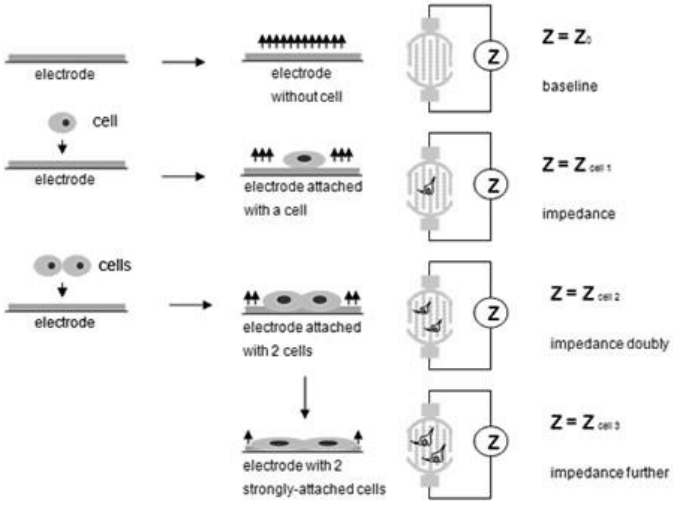


Figure 106 iCELLigence principle
The system is based on 3 components: an electronic analyzer inside the cell culture incubator, an iPad able to run the needed software and electronic plates called E plates L8 with microelectrodes. The status of the cells is reflected by the changes on the electrical impedance, defined as the opposition of a system to the current if a voltage is applied: in presence of cells the electrode impedance is increased. The quantity of cells that interact and the kind of this interaction to the electrode influence the level of impedance. (http://www.aceabio.com)

V.5.2.2 Results

V.5.2.2.1 Properties of cell models used to investigate functional effects of PCDHs expression

We established four stable cell lines expressing, respectively, EYFP tagged PCDH11X, PCDH11Y, or PCDH10 and also a mock cell line expressing EYFP alone. We used the Doxycycline-inducible Flp-in™ T-Rex™ HEK293T cell line (Life Technologies) to generate these stable cell lines because it allows single copy integration of the gene target. The Flp-in™ T-Rex™ host cell line contains a single FRT site and expresses the Tet repressor. More details about the Flp-In and the T-rex Systems can be found in the methodology section. As a result, the Flp-In™ T-Rex™ expression cell lines should exhibit the following phenotype: hygromycin resistance, zeocin™ sensitivity, lack of ß-galactosidase activity, blasticidin resistance and tetracycline-regulated expression of the gene of interest. Foremost, the constructs were produced by cloning into the pcDNATM5/FRT plasmid different full-length PCDH cDNAs, each time fused in frame to EYFP: PCDH11X-EYFP, PCDH11Y-EYFP, and PCDH10-EYFP. Moreover, one construct contained EYFP alone (mock control). Each construct encodes also a hygromycin resistance gene for selection of positive transfectants. After successful Flp-mediated single copy integration, expression could be induced by Doxycycline. The stable cell line for PCDH11X-EYFP and the mock cell line were the first ones to be established and the expression was checked with and without induction (**Figure 107** and **Figure 108**).

PCDH10 was expressed as well as a EYFP fusion protein in the Flp-In TRex HEK293T cells but also in non- inducible HeLa cells (**Figure 109**). The PCDH10-EYFP stable cell line was used to confirm the results obtained using the Flp-In TRex HEK293T PCDH10-EYFP cell line. After preliminary tests we decided to use only Flp-In TRex HEK293T PCDH10-EYFP and Flp-In TRex HEK293T PCDH11X-EYFP (with and without DOX) because of a stronger expression level.

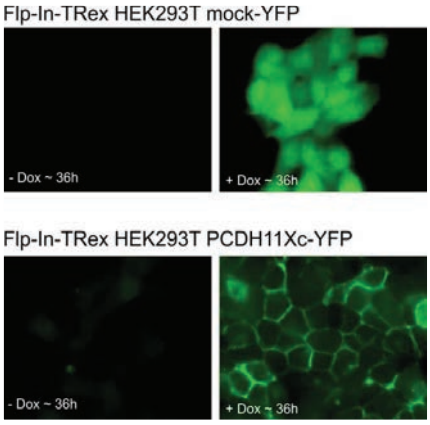


Figure 107 Tetracycline inducible cell lines.
Two cell lines expressing either EYFP alone (mock control, top panel) or PCDH11X-EYFP (lower panel) after induction by doxycycline, a clear plasma membrane staining can be observed for PCDH11X, whereas the mock control shows cytoplasmic staining (Uta Brunner, unpublished data).

Flp-In-TRex Hek293T PCDH11Y-YFP

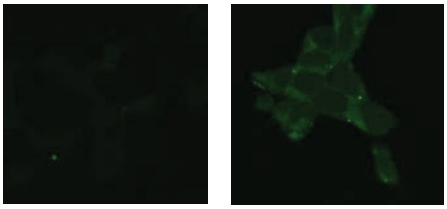


Figure 108 Tetracycline inducible cell lines.
Stable cell line expressing PCDH11Y-EYFP with or without induction by Doxycycline (36 h). Compared with the stable cell line for PCDH11X we obtained a very low expression level, despite correct protein localization.

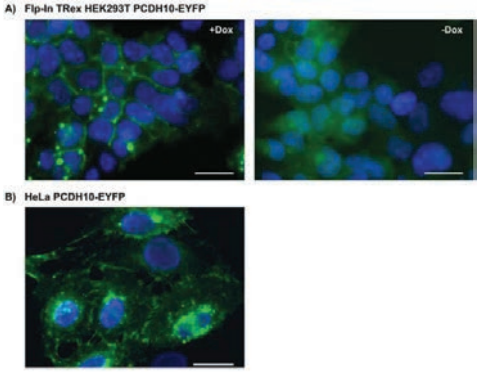


Figure 109 PCDH10-EYFP expressing stable cell lines
A) HEK293T stably expressing PCDH10-EYFP. Before fixations cells were induced with Doxycycline for 24h (+Dox) or left untreated (-Dox). PCDH10 was observed at the plasma membrane. B) HeLa cells stably expressing PCDH10-EYFP generated with lentiviral transduction. PCDH 10 was enriched at cell-cell contacts and filopodia protrusions (I.Kahr, personal communication)

V.5.2.2.2 Effect of PCDH expression on cell migration velocity

In a 2D-cell exclusion zone assay[77,79] and using time lapse microscopy, we compared the migration kinetics of the stable HEK293T-FlpInTRex cell line in which PCDH10-EYFP expression was induced using a 48 h doxycycline treatment, to that of uninduced cells, used as control. **Figure 110A** shows the extent of migration at selected time points based on the phase contrast images (top and middle panels) and also demonstrates the expression level of the fluorescent fusion protein in the induced cell population (**Figure 110A,** bottom panels). The evolution of the area covered by the confluent cell layer in time, as it migrates in the central free zone in normal growth medium (10% FBS), is plotted for both conditions in **Figure 110B**. The cells with PCDH10 migrated slightly, but significantly faster than the control cells ($p<0.001$, **Figure 110C**). The median migration velocity is derived from the slope of the linear part of the curve (0-25h). In these cells, the positive effect of PCDH10 was not observed from the in conditions in which the migration velocity is inherently lower: lower serum condition (1% FBS) or 1%FBS in combination with 5nM EGF.

Based on the same assay, we showed that cells of a HeLa-derived cell line stably expressing a PCDH10-mKate2 fusion protein, also migrate significantly faster than cells of a control cell line (expressing only mKate2), both in 10 % serum and in 1% serum, 5nM EGF (**Figure 111**).

A stable HEK293T-FlpInTRex cell line, in which PCDH11X-EYFP is induced by doxycycline treatment, was also tested. In contrast to the effect of PCDH10 described above, a change in migration properties was not observed for cells expressing PCDH11x at any of the tested conditions (**Figure 112**). This may indicate specific differences between the two family members; however, we need to consider that the PCDH11X expression level seems to be lower than PCDH10 and this may also contribute to this difference. Further analyses are required.

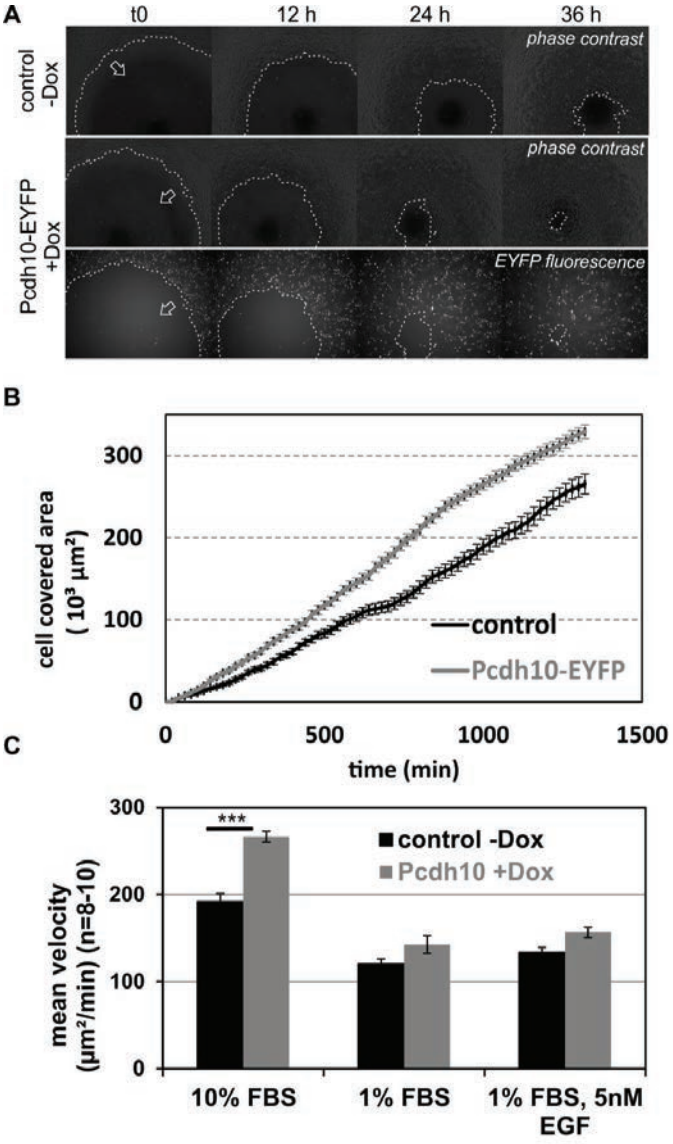


Figure 110 Effect of PCDH10-EYFP expression on migration of inducible FlpIn-TRex HEK293T stable cell line. *A. Images selected from time lapse sequence without (top) or with (middle) expression of PCDH10. The arrows indicate the direction of migration of the confluent cell layer in the central cell free zone; the dotted line delineates the cell layer and the cell-free zone. Bottom and middle panels: corresponding phase contrast and EYFP images. B. Mean cell covered area in time for control cells (9 replicate samples) and PCDH 10 cells (10 replicates) in 10% FBS containing medium; error bars are SEM. C Mean velocity derived from plots as in B for control and PCDH10 cells tested under different conditions. Error bar is SEM, ***: p<0.01; n, number of replicates within experiment.*

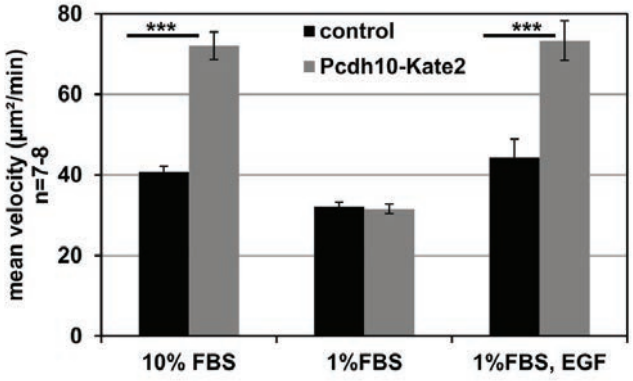


Figure 111 Effect of PCDH-10-MKate2 expression on migration of stable HeLa cell line.
*Mean cell velocity derived covered area in time for control cells (mKate2) and PCDH 10-mKate2 cells under different conditions; error bars are SEM. ***: p<0.01; n number of replicates within experiment.*

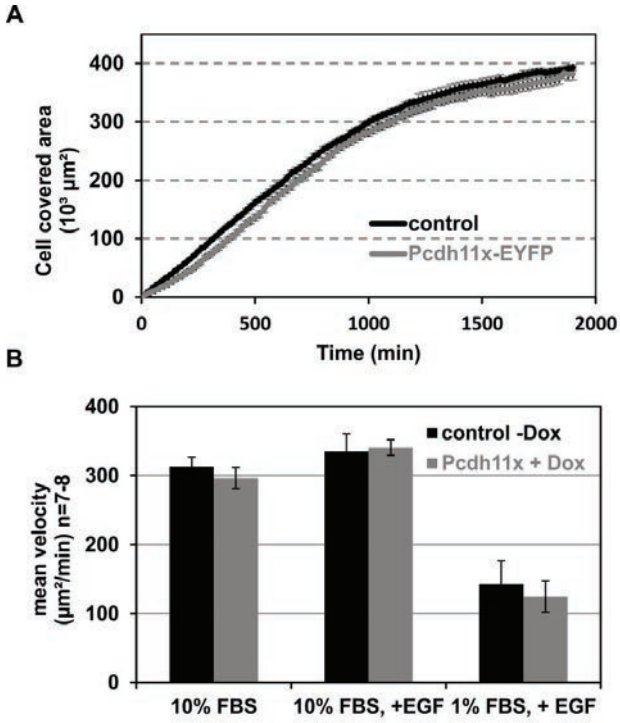


Figure 112 Effect of PCDH-11x-EYFP expression on migration of inducible HEK-FlpIn-TRex stable cell line.
A. Mean cell covered area in time for control cells (8 replicate samples) and PCDH 11x cells (7 replicates) in 10% FBS containing medium; error bars are SEM. B. Mean velocity derived from plots as in B for control and PCDH 11x cells tested under different conditions. Error bar is SEM, n number of replicates within experiment.

V.5.2.2.3 Effect of PCDH expression on cell growth

The proliferation of the HEK293T-FlpInTRex cell lines with induced expression of either EYFP-tagged PCDH10 or PCDH11X was compared to that of the respective non-induced cells. Cell number was scored based on the level of metabolic activity (using XTT- proliferation assay producing a soluble colored formazan[80] (see methods). Beforehand, the linearity of the readout was tested in absence and presence of doxycycline by measuring formazan-based absorption for a cell number range of 750-40000 cells per well (**Figure 113A**). The effect of different PCDHs expression on cell growth was measured (using a starting cell number of 4500 cells/well) at time zero and at three later time points with a maximum of 2.5 days (60h). The signal at each time point versus that at time zero is indicative of the fold change in proliferation under the different conditions. We note that at 60 h the signal slightly exceeded the maximal signal in **Figure 113A** suggesting the signal ratios at this time point could be an underestimation.

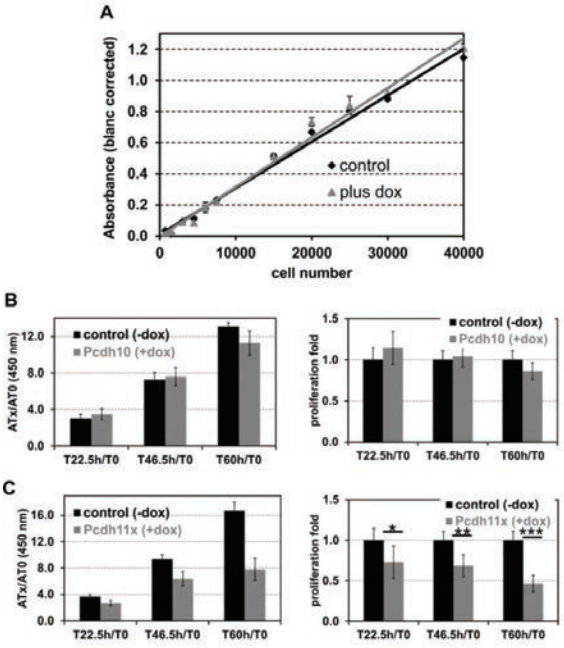


Figure 113 Effect of PCDH10 or PCDH11X-EYFP expression on proliferation of inducible HEK293t stable cell lines using XTT-assay.

*A. Linear response of signal for the cell number range 0-40.000 in absence or presence of doxycycline (dox). B, C. (left) Blanc corrected average absorption (A) ratio versus time zero (T0) at the time points indicated. (right) Average proliferation fold for the cells expressing PCDH10 (B) or PCDH11x (C) versus control cells (i.e. the corresponding cells without doxycycline treatment). In C, the control condition is set to 1 for each time point. Error bars are SEM; n=6; *, **, *** p < 0.1, 0.05 and 0.01 respectively.*

Figure 113B shows that the signal ratio (left) at the different time points of the HEK293T-FlpInTRex PCDH10-EYFP cells and the corresponding fold change in proliferation (with ratio of control cells set to 1, right panel) are not significantly different from that of the control, at least at the induced level of PCDH10 and within the tested time period. In contrast, the presence of PCDH11X-EYFP, at the induced level, altered the ratio of the measured absorption values (left panel) and thus reduced the fold change of proliferation of the HEK293Tcells (right panel, control cells set to 1) by ~ 30 % after 48 h to~50% at 60 h (**Figure 113C**) This effect on cell growth needs to be confirmed by repeating the current experiments and by using a complementary proliferation assay.

V.5.2.2.4 Effect of expression of different PCDHs on cell adhesion on collagen

Cells display integrin-dependent attachment to and spreading on the extracellular matrix protein collagen. We used the an impedance-based real-time cell analysis system (iCELLigence), which allows for label-free, dynamic monitoring of these cellular properties[81]. As described in the methods section, the impedance is expressed as Cell Index and the change in impedance ΔCI in a given time interval ΔT is used to compare adhesion and spreading kinetics between conditions.

We compared the adhesion and spreading kinetics of HEK293T-FlpInTRex cells without or with induced expression of either EYFP-tagged PCDH10 or PCDH11X. As for the effects on migration and proliferation, the observed effect after induction is also determined by the level of expression of each PCDH.

The comparison between the four conditions was independently performed three times up till now. Within each experiment, we evaluated the difference between the –Dox and + Dox condition. Interpretation across experiments was hampered by the fact the rate in CI increase and the plateau CI-value varies between biological replicates. We here present data of one experiment (**Figure 114**). This shows no effect of PCDH10 expression (at the induced level) and a faster and/or more extensive spreading of the cells expressing PCDH11X at the induced level. This was consistent over two and three biological replicate experiments for PCDH10 and PCDH 11x, respectively. We note, however, that the differences induced by PCDH expression at the tested expression level are relatively small. **Figure 114A, B** shows the kinetics (CI versus time); **Figure 114C, D** the adhesion-spreading rates in different time intervals: 10- 30 minutes; 30-60 and 10-60 minutes. For the cells expressing PCDH11X the increased rate is mainly apparent early, in the first time interval. Additional experiments, preferably using imaging, need to confirm this and determine to which extend the cells already start spreading in this time interval.

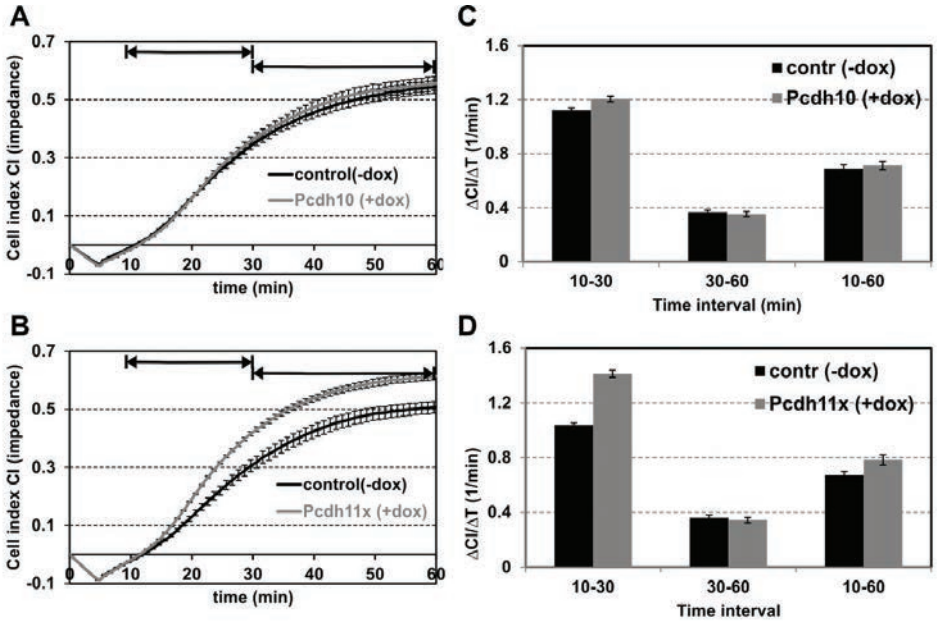


Figure 114 Effect of PCDH10-EYFP or PCDH11X-EYFP expression on adhesion and spreading kinetics of inducible HEK-FlpIn-TRex stable cell lines.

A, B. *Real time monitoring of the impedance (expressed as cell index) increase upon attachment of a same amount of cells on the well surface of E8-plates for uninduced (control) and doxycycline-induced cells. Error is SEM, n=4 per condition. The time intervals used for graphs in C, D are indicated. At time 0, before the addition of cells, the signal in all cells is set to zero. The data shown are from one experiment and the observed relative effect was also observed in two and three independent experiments (out of three performed) for PCDH 10 and 11x, respectively.* ***C, D*** *Change of cell index CI in the indicated time intervals for the tested conditions. Error is SD, n=4. Note the increase in adhesion-spreading rate in the 10-30 minute interval for the protocadherin 11x expressing cells compared to control cells (**D**). C, D are based on the data of the experiment shown in A, B.*

V.5.2.3 Conclusions

We explored cellular effects of expressing either PCDH10 or PCDH11X at a given level in a set of stable and/or inducible cell lines. The major goal was not so much to compare the differential effect of the different PCDHs but to establish *in vitro* cell based assays in which the PCDH10 or PCDH11X expression results in an effect that can be used in subsequent assays to test the effect of interactions with the candidate molecular partners that we identified.

In summary we demonstrated that PCDH10 expressing cells displayed a moderate increase in cell migration speed on a 2D collagen-coated surface, but no effect on cell proliferation and no

consistent effect on the rate of cell attachment or spreading on collagen type I. We noted that the positive effect on cell migration speed was observed both in the inducible PCDH10-expressing HEKk293T cell line and in the stably transfected HeLa cell line. The PCDH11X expressing HEK293T cells displayed no significant effect on migration speed. However, the presence of PCDH11X significantly induced slower growth in vitro and seemed to induce the cells to attach more efficiently to and/or spread more quickly on a collagen coated surface. The effect of PCDH10 on cell migration and of PCDH11X on other characteristics are substantial and open possibilities to use these cell based effects for functionally validating a number of the partners identified previously using interactomics.

Upregulation observed in cell migration in presence of PCDH10 is in contrast with the tumor suppressor role proposed for PCDH10 but can correlate with the recruitment of NAP1 to the cell membrane by interaction of PCDH10 and the WAVE complex (see below).

For further investigations on migration, proliferation and adhesion using the assay we established will be important to consider the effect in different cell lines and to compare normalized expression level values in transiently and stable transfected cells.

V.5.3 WAVE complex and cell migration

Actin polymerization is responsible for the movement of the cells and is regulated by nucleation promoting factors such as WASP, WAVE and Ena/VASP. Different δPCDHs have been shown to play a role in the regulation of this delicate pathway by interacting with proteins of the complex. To elucidate better the mechanisms exploited by Pcdhs, we performed binary MAPPIT experiments on candidate PCDH interactors, previously indicated by Array MAPPIT. The results obtained could confirm some of these candidates. So, KLF4 (Kruppel-like factor 4) was identified as a confirmed interaction partner for PCDH9 and PCDH19, and ARL13B as an interaction partner for PCDH11Y, PCDH18 and PCDH19. KLF4 is a member of the KLF family of transcription factors and regulates proliferation, differentiation, apoptosis and somatic cell reprogramming. Interaction between PCDHs and NAP1 is indirect and cannot be shown with MAPPIT. As previously explained, Sra1 and Abi2 form an interface where PCDHs can bind in the WRC and the interaction cause the formation of the penta complex. This makes this interaction not suitable for this technique. It has been shown that (in apparent contradiction with its suggested tumor suppressor activity) PCDH10 enhances cell migration in astrocytoma cell lines as they recruit the WAVE complex from the lamellipodia to cell-cell contacts. This may alter the F-actin organization leading to N-cadherin redistribution. In this way,

N-cadherin loses its ability to induce contact inhibition, resulting in an uncooperative, accelerated cell movement[60]. Recently, a BioID experiment (unpublished data of van Roy's lab) identified some of my MAPPIT hits as proximity partners for PCDH10: BAIAP2 (candidate interactor of PCDH9 and PCDH11X), CORO1B (candidate interactor of PCDH11X), DPYSL5 (candidate interactor of PCDH9), SH3D19 (candidate interactor of PCDH9) and TTLL1 (candidate interactor of PCDH11X).

Our experiments have shown that PCDH10 expressing HEK293 and HeLa cells display a moderate increase in cell migration speed. Currently, plasmids encoding the candidate PCDH10 interactors identified by MAPPIT and BioID experiments are transfected in PCDH10 and PCDH11X expressing HEK293 cell line derivatives. These transfectants are then studied in migration experiments to elucidate if the interaction with PCDHs can influence the cell migration speed.

V.6 References

1. Streit, S., Michalski, C. W., Erkan, M., Kleeff, J. & Friess, H. Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nat. Protoc.* **4,** 37–43 (2009).

2. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3,** RESEARCH0034 (2002).

3. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17,** 10 (2011).

4. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–11 (2009).

5. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28,** 511–5 (2010).

6. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–2 (2010).

7. Risso, D. *et al.* GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics* **12,** 480 (2011).

8. Kim, J., Lee, J. & Iyer, V. R. Global identification of Myc target genes reveals its direct role in mitochondrial biogenesis and its E-box usage in vivo. *PLoS One* **3,** e1798 (2008).

9. Anderson, P. D. *et al.* Nkx3.1 and Myc crossregulate shared target genes in mouse and human prostate tumorigenesis. *J. Clin. Invest.* **122,** 1907–19 (2012).

10. Barisone, G. A., Yun, J.-S. & Díaz, E. From cerebellar proliferation to tumorigenesis: new insights into the role of Mad3. *Cell Cycle* **7,** 423–7 (2008).

11. Reyes-Turcu, F. E., Ventii, K. H. & Wilkinson, K. D. Regulation and cellular roles of ubiquitin-specific deubiquitinating enzymes. *Annu. Rev. Biochem.* **78,** 363–97 (2009).

12. Ohta, T. & Fukuda, M. Ubiquitin and breast cancer. *Oncogene* **23,** 2079–2088 (2004).

13. Pal, A. & Donato, N. J. Ubiquitin-specific proteases as therapeutic targets for the treatment of breast cancer. *Breast Cancer Res.* **16,** 461 (2014).

14. Orlowski, R. Z. & Dees, E. C. The role of the ubiquitination-proteasome pathway in breast cancer: applying drugs that affect the ubiquitin-proteasome pathway to the therapy of breast cancer. *Breast Cancer Res.* **5,** 1–7 (2003).

15. Jin, H. & Zangar, R. C. Protein modifications as potential biomarkers in breast cancer. *Biomark. Insights* **4,** 191–200 (2009).

16. Hanash, S. M., Pitteri, S. J. & Faca, V. M. Mining the plasma proteome for cancer biomarkers. *Nature* **452,** 571–9 (2008).

17. Jensen, O. N. Interpreting the protein language using proteomics. *Nat. Rev. Mol. Cell Biol.* **7,** 391–403 (2006).

18. Spickett, C. M., Pitt, A. R., Morrice, N. & Kolch, W. Proteomic analysis of phosphorylation, oxidation and nitrosylation in signal transduction. *Biochim. Biophys. Acta* **1764,** 1823–41 (2006).

19. Ouyang, X. *et al.* Association of ErbB2 Ser1113 phosphorylation with epidermal growth factor receptor co-expression and poor prognosis in human breast cancer. *Mol. Cell. Biochem.* **218,** 47–54 (2001).

20. Vazquez-Martin, A., Oliveras-Ferraros, C., Colomer, R., Brunet, J. & Menendez, J. A. Low-scale phosphoproteome analyses identify the mTOR effector p70 S6 kinase 1 as a specific biomarker of the dual-HER1/HER2 tyrosine kinase inhibitor lapatinib (Tykerb) in human breast carcinoma cells. *Ann. Oncol.* **19,** 1097–109 (2008).

21. Jung, C.-R. *et al.* Enigma negatively regulates p53 through MDM2 and promotes tumor cell survival in mice. *J. Clin. Invest.* **120,** 4493–506 (2010).

22. Kales, S. C., Nau, M. M., Merchant, A. S. & Lipkowitz, S. Enigma Prevents Cbl-c-Mediated Ubiquitination and Degradation of RETMEN2A. *PLoS One* **9,** e87116 (2014).

23. Durick, K., Gill, G. N. & Taylor, S. S. Shc and Enigma are both required for mitogenic signaling by Ret/ptc2. *Mol. Cell. Biol.* **18,** 2298–308 (1998).

24. Schalm, S. S., Ballif, B. A., Buchanan, S. M., Phillips, G. R. & Maniatis, T. Phosphorylation of protocadherin proteins by the receptor tyrosine kinase Ret. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 13894–9 (2010).

25. Zhang, W. *et al.* Four and a half LIM protein 2 (FHL2) negatively regulates the transcription of E-cadherin through interaction with Snail1. *Eur. J. Cancer* **47,** 121–30 (2011).

26. Gabay, M., Li, Y. & Felsher, D. W. MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. *Cold Spring Harb. Perspect. Med.* **4,** a014241–a014241 (2014).

27. Qi, D., Jin, H., Lilja, T. & Mannervik, M. Drosophila Reptin and other TIP60 complex components promote generation of silent chromatin. *Genetics* **174,** 241–51 (2006).

28. Kanemaki, M. *et al.* TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* **274,** 22437–44 (1999).

29. Makino, Y., Kanemaki, M., Kurokawa, Y., Koji, T. & Tamura, T. a. A rat RuvB-like protein, TIP49a, is a germ cell-enriched novel DNA helicase. *J. Biol. Chem.* **274,** 15329–35 (1999).

30. Frank, S. R. *et al.* MYC recruits the TIP60 histone acetyltransferase complex to chromatin. *EMBO Rep.* **4,** 575–80 (2003).

31. Fang, S. *et al.* Silencing of PCDH10 in hepatocellular carcinoma via de novo DNA methylation independent of HBV infection or HBX expression. *Clin. Exp. Med.* **13,** 127–34 (2013).

32. Ying, J. *et al.* Frequent epigenetic silencing of protocadherin 10 by methylation in multiple haematologic malignancies. *Br. J. Haematol.* **136,** 829–32 (2007).

33. Li, Z. *et al.* Role of PCDH10 and its hypermethylation in human gastric cancer. *Biochim. Biophys. Acta* **1823,** 298–305 (2012).

34. Li, Z. *et al.* Epigenetic inactivation of PCDH10 in human prostate cancer cell lines. *Cell Biol. Int.*

35, 671–6 (2011).

35. Yu, J. *et al.* Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* **136,** 640–51.e1 (2009).

36. Hough, C. D., Cho, K. R., Zonderman, A. B., Schwartz, D. R. & Morin, P. J. Coordinately up-regulated genes in ovarian cancer. *Cancer Res.* **61,** 3869–76 (2001).

37. Costea, D. E. *et al.* Identification of two distinct carcinoma-associated fibroblast subtypes with differential tumor-promoting abilities in oral squamous cell carcinoma. *Cancer Res.* **73,** 3888–901 (2013).

38. Miles, R. R., Crockett, D. K., Lim, M. S. & Elenitoba-Johnson, K. S. J. Analysis of BCL6-interacting proteins by tandem mass spectrometry. *Mol. Cell. Proteomics* **4,** 1898–909 (2005).

39. Cascione, L. *et al.* Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer. *PLoS One* **8,** e55910 (2013).

40. Seshagiri, S. *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* **488,** 660–664 (2012).

41. Dulak, A. M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45,** 478–86 (2013).

42. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44,** 325–40 (2011).

43. Grando, S. A. Connections of nicotine to cancer. *Nat. Rev. Cancer* **14,** 419–29 (2014).

44. Schaal, C. & Chellappan, S. P. Nicotine-mediated cell proliferation and tumor progression in smoking-related cancers. *Mol. Cancer Res.* **12,** 14–23 (2014).

45. Yu, C.-S., Chen, Y.-C., Lu, C.-H. & Hwang, J.-K. Prediction of protein subcellular localization. *Proteins* **64,** 643–51 (2006).

46. Yu, C.-S., Lin, C.-J. & Hwang, J.-K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* **13,** 1402–6 (2004).

47. Hoglund, A., Donnes, P., Blum, T., Adolph, H.-W. & Kohlbacher, O. MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* **22,** 1158–1165 (2006).

48. Horton, P. *et al.* WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* **35,** W585–7 (2007).

49. Shen, H.-B. & Chou, K.-C. Hum-mPLoc: an ensemble classifier for large-scale human protein subcellular location prediction by incorporating samples with multiple sites. *Biochem. Biophys. Res. Commun.* **355,** 1006–11 (2007).

50. Nguyen Ba, A. N., Pogoutse, A., Provart, N. & Moses, A. M. NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics* **10,** 202 (2009).

51. Mooney, C., Wang, Y.-H. & Pollastri, G. SCLpred: protein subcellular localization prediction by N-to-1 neural networks. *Bioinformatics* **27,** 2812–9 (2011).

52. Pierleoni, A., Martelli, P. L., Fariselli, P. & Casadio, R. BaCelLo: a balanced subcellular localization predictor. *Bioinformatics* **22,** e408–16 (2006).

53. Boulikas, T. Nuclear localization signals (NLS). *Crit. Rev. Eukaryot. Gene Expr.* **3,** 193–227 (1993).

54. Lange, A. *et al.* Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.* **282,** 5101–5 (2007).

55. Heggem, M. A. & Bradley, R. S. The cytoplasmic domain of Xenopus NF-protocadherin interacts with TAF1/set. *Dev. Cell* **4,** 419–29 (2003).

56. Bonn, S., Seeburg, P. H. & Schwarz, M. K. Combinatorial expression of alpha- and gamma-protocadherins alters their presenilin-dependent processing. *Mol. Cell. Biol.* **27,** 4121–32 (2007).

57. Tai, K., Kubota, M., Shiono, K., Tokutsu, H. & Suzuki, S. T. Adhesion properties and retinofugal expression of chicken protocadherin-19. *Brain Res.* **1344,** 13–24 (2010).

58. Hayashi, S. *et al.* Protocadherin-17 Mediates Collective Axon Extension by Recruiting Actin Regulator Complexes to Interaxonal Contacts. *Dev. Cell* **30,** 673–687 (2014).

59. Hayashi, S. & Takeichi, M. Emerging roles of protocadherins: from self-avoidance to enhancement of motility. *J. Cell Sci.* **128,** 1455–1464 (2015).

60. Nakao, S., Platek, A., Hirano, S. & Takeichi, M. Contact-dependent promotion of cell migration by the OL-protocadherin-Nap1 interaction. *J. Cell Biol.* **182,** 395–410 (2008).

61. Biswas, S. *et al.* Protocadherin-18b interacts with Nap1 to control motor axon growth and arborization in zebrafish. *Mol. Biol. Cell* **25,** 633–42 (2014).

62. Stoeckli, E. T. Protocadherins: not just neuron glue, more too! *Dev. Cell* **30,** 643–4 (2014).

63. Aamar, E. & Dawid, I. B. Protocadherin-18a has a role in cell adhesion, behavior and migration in zebrafish development. *Dev. Biol.* **318,** 335–346 (2008).

64. Chen, B. *et al.* The WAVE regulatory complex links diverse receptors to the actin cytoskeleton. *Cell* **156,** 195–207 (2014).

65. Eden, S., Rohatgi, R., Podtelejnikov, A. V, Mann, M. & Kirschner, M. W. Mechanism of regulation of WAVE1-induced actin nucleation by Rac1 and Nck. *Nature* **418,** 790–3 (2002).

66. Krause, M. & Gautreau, A. Steering cell migration: lamellipodium dynamics and the regulation of directional persistence. *Nat. Rev. Mol. Cell Biol.* **15,** 577–90 (2014).

67. Law, A.-L. *et al.* Lamellipodin and the Scar/WAVE complex cooperate to promote cell migration in vivo. *J. Cell Biol.* **203,** 673–89 (2013).

68. Hultquist, A. *et al.* Mad 1 inhibits cell growth and proliferation but does not promote differentiation or overall survival in human U-937 monoblasts. *Mol. Cancer Res.* **2,** 464–76 (2004).

69. Ding, L. *et al.* Human four-and-a-half LIM family members suppress tumor cell growth through a TGF-beta-like signaling pathway. *J. Clin. Invest.* **119,** 349–61 (2009).

70. Müller, J. M. *et al.* FHL2, a novel tissue-specific coactivator of the androgen receptor. *EMBO J.* **19,** 359–69 (2000).

71. Hill, A. A. & Riley, P. R. Differential regulation of Hand1 homodimer and Hand1-E12 heterodimer activity by the cofactor FHL2. *Mol. Cell. Biol.* **24,** 9835–47 (2004).

72. Labalette, C., Renard, C.-A., Neuveut, C., Buendia, M.-A. & Wei, Y. Interaction and functional cooperation between the LIM protein FHL2, CBP/p300, and beta-catenin. *Mol. Cell. Biol.* **24,** 10689–702 (2004).

73. Samson, T. *et al.* The LIM-only proteins FHL2 and FHL3 interact with alpha- and beta-subunits of the muscle alpha7beta1 integrin receptor. *J. Biol. Chem.* **279,** 28641–52 (2004).

74. Martin, B. *et al.* The LIM-only protein FHL2 interacts with beta-catenin and promotes differentiation of mouse myoblasts. *J. Cell Biol.* **159,** 113–22 (2002).

75. Coghill, I. D. *et al.* FHL3 is an actin-binding protein that regulates alpha-actinin-mediated actin bundling: FHL3 localizes to actin stress fibers and enhances cell spreading and stress fiber disassembly. *J. Biol. Chem.* **278,** 24139–52 (2003).

76. Han, W. *et al.* FHL2 interacts with and acts as a functional repressor of Id2 in human neuroblastoma cells. *Nucleic Acids Res.* **37,** 3996–4009 (2009).

77. HUYCK, L. Cancer cell migration and proliferation: Advanced quantitative in vitro analysis in threedimensional matrices at higher throughput and focus on ADF/cofilin regulatory networks. (2012). at <https://biblio.ugent.be/record/4337122>

78. Masuzzo, P. *et al.* CellMissy: a tool for management, storage and analysis of cell migration data produced in wound healing-like assays. *Bioinformatics* **29,** 2661–3 (2013).

79. Hulkower, K. I. & Herber, R. L. Cell migration and invasion assays as tools for drug discovery. *Pharmaceutics* **3,** 107–24 (2011).

80. Scudiero, D. A. *et al.* Evaluation of a soluble tetrazolium/formazan assay for cell growth and drug sensitivity in culture using human and other tumor cell lines. *Cancer Res.* **48,** 4827–33 (1988).

81. Atienza, J. M., Zhu, J., Wang, X., Xu, X. & Abassi, Y. Dynamic monitoring of cell adhesion and spreading on microelectronic sensor arrays. *J. Biomol. Screen.* **10,** 795–805 (2005).

Conclusions

Protein-protein interactions are an essential part of most of biological processes and due to this relevance more and more methods to study them have been developed in the decades. Different methods offer different solutions: some can detect the interaction between two known proteins while others are suitable to identify a list of previously unknown interactors for a protein of interest. An overview of genetic and biochemical strategies is given in Chapter II of this thesis.

The complete human proteome has been estimated to include about 400.000 protein-protein interactions (PPIs), indicating the complexity of this network; unravelling this complexity may provide answers to many unsolved questions in biology. The studies described in this Doctorate dissertation focused on δ-protocadherins (δPcdhs), due to their relevance in many human biological events, both healthy and pathologic i.e. in many types of cancer[1]. Only a very limited number of studies describing δPcdhs interaction partners have been published to date. Protein phosphatase 1 α (PP1α) is a well-known interactor of the CM3 domain present in δ1Pcdhs[2]. Moreover, different δPcdhs have been observed to interact with the WASP-family verprolin homologous protein (WAVE) complex, confirming that these Pcdh family members share common binding mechanisms[3,4]. Furthermore, several studies have shown that δPcdhs can bind, directly or indirectly, to classical cadherins and influence their adhesive behaviour[5–7].

This dissertation investigates the δPcdh interactome and studies the modulation of δPcdh functions by the candidate interaction partners.

In order to identify new intracellular interaction partners of the δPcdh family members, we carried out PPI screenings in yeast (yeast two-hybrid screen, Y2H) and in mammalian cells for different members of the family. We encountered limitations in the yeast system and in the first attempts to use a cDNA library screening on mammalian cells. We speculated that the Y2H screens failed because PCDHs and/or the prey clones were either unable to fold properly or were missing important post-translational modifications. For the MAPPIT (mammalian protein-protein interaction trap) cDNA library, limitations occurred because clones used in this technique may not represent the full length DNAs, causing incorrect folding, or because cDNA sub-domains might be cloned in the wrong frame. To overcome these limitations and provide a system closer to physiological conditions, the Array MAPPIT[8] screening was developed and it was selected in this study as the most useful screening method for intracellular interaction partners of δPcdhs. Different elements in the Array MAPPIT procedure cooperate to create a near-optimal physiological context in which PCDHs and putative preys may interact, including:

- MAPPIT operates in mammalian cells
- The PCDH cytoplasmic tail is fused to the C-terminus of a membrane receptor: the interaction can occur in the natural cell compartment.
- The physical separation of interactor and effector zone: the interactions occur at the membrane, while the readout is in the nucleus. This avoids interference of the hybrid proteins with the reporter activity.
- An extra control is given by the signal being ligand-dependent.

Taken together these characteristics indicate the Array MAPPIT a valid assay to identify interaction partners of PCDH family members. Nonetheless, some of our baits, especially PCDH10, showed a high background which makes the analysis less objective and precise than usual. A collection of 10,000 open reading frames (ORFs) was used as a prey library to identify novel interaction partners. Confirmation using binary MAPPIT assays was performed with the aim to narrow down the long list of candidate interactors and to increase the stringency of the selection. In order to identify common intracellular binding partners among the δPcdhs tested we performed cross-tests. In this way, we obtained a priority list of candidate interactors, which needed independent confirmation. Next to MAPPIT-specified candidates, we considered other interaction candidates as suggested by literature studies. A priority list was made following an Ingenuity Pathway Analysis (IPA) to look for partners belonging to the same pathway, literature indications and our interest for specific or common cancer-related candidates. Our priority list for biochemical studies included myc-associated factor X (MAX), MAX dimerization protein 3 (MXD3) and MAX interactor 1 (MXI1), all three belonging to the same pathway; Cullin 5(CUL5), Four And A Half LIM Domains 3 (FHL3), PDZ And LIM Domain 7 (Enigma, PDLIM7), Doublesex and mab-3 related transcription factor-like C 1 (DMRTC1) and tripartite motif-containing 23 (TRIM 23). Moreover, the MAPPIT data revealed other interesting candidates, including Proteasome 26S Subunit, ATPase 1 (PSMC1), ubiquitin-specific proteases (USP) family members and actin-related proteins.

In this thesis we show that MAX interacts with PCDH1, PCDH7, PCDH9, PCDH10 and PCDH11X, while interaction with MXD3 could be confirmed in CoIP only for PCDH10. MAX is central in a transcription factor network that closely relates to the MYC family of oncoproteins and to a group of MYC antagonists (the MAD family). The MAD family has been demonstrated to have several different functions related to terminal differentiation, inhibition of cell cycle progression and tumor suppression[9–19]. We hypothesize that the interaction with PCDHs could boost the physiological functions of the MAD family components. To investigate this speculation *in vitro*, we performed a MYC reporter assay in HEK293T cells stably expressing PCDH10 or PCDH11X to monitor whether the

Myc-regulated signaling pathway activities in cells are influenced by the co-expression of PCDHs. The results, however, showed no differences between induced and non- induced HEK293T cells in case of expression of either PCDH. An interesting future follow-up experiment could involve a p53-responsive construct in a similar set-up to test the effects of PCDH over-expression or silencing on the influence of Enigma/PDLIM7 in the context of the p53 pathway. Enigma/PDLIM7 is involved in neuronal signaling and tumor cell growth; it associates with actin filaments in fibroblasts via its PDZ domain and it has been shown to negatively regulate p53 through inhibition of MDM2 self-ubiquitination[20–22]. The p53 protein has been shown to be destabilized by stress factors like DNA-damaging agents (e.g., doxorubicin) due to a MDM2 binding-mediated degradation. The function of MDM2 is positively regulated by MDM4 (MDMX) and PDLIM7. Studies under doxorubicin treatment might lead to a better understanding of the mechanisms involved in the PDLIM7-dependent regulation of the p53 pathway. This evidence makes PDLIM7 an interesting candidate to understand PCDH functions. Finally, two more candidate interactors of PCDH10 and PCDH11X , FHL2 and FHL3, have also been shown to play a role in cancer as oncogenes and are involved in cell migration[23–28]. We were able to demonstrate that both PCDH10 and PCDH11X interact with FHL2.

Migration, proliferation and adhesion are three critical mechanisms playing important roles in development and cancer. We established in vitro cell based assays in order to investigate the role of PCDHs, alone or influenced by the putative interaction partners identified in our study. The effect of PCDH expression can be used in subsequent assays to investigate the roles played by the candidate molecular interaction partners identified in our experiments. A cell exclusion zone migration assay was established using inducible HEK293T cells expressing enhanced yellow fluorescent protein (EYFP)-tagged PCDH10 or PCD11x proteins. Our experiments show that PCDH10, but not PCDH11x, induces faster migration in HEK293T cells. We confirmed this observation in a PCDH10-EYFP expressing HeLa-derived cell line. Interestingly, this result is in apparent contrast with the tumor suppressor role proposed for PCDH10. PCDH10 has been shown to interact with the WAVE complex through a specific surface of the fully assembled wave regulatory complex (WRC), composed by Steroid Receptor RNA Activator 1 (SRA1) and Abl-Interactor 2 (ABI2)[29]. This interaction induces the recruitment of NAP1 at the membrane with alteration of F-actin organization and a N-cadherin redistribution. This results in an accelerated cell movement and can be an explanation for our observations. Both the HEK293T-FlpInTRex PCDH10-EYFP and the Hela-PCDH10-mKate2 cells express a long isoform of PCDH10. It might be interesting, in future experiments, to generate also similar cell lines expressing only the short isoform to investigate whether the latter plays a different role in the cells or has different interactions with specific candidates: an interaction partner influencing the

behavior of a cell line expressing the long isoform but not of a cell line expressing the short isoform could indicate specificity, especially in case the interactor is also found to bind to the cytoplasmic domain present only in the long isoform. In this context, MAPPIT experiments are planned for the short isoform of PCDH10. Our studies are at the moment focused on Pcdh10 because of specific reasons: PCDH10 exists only in two isoforms in humans, contrary to the numerous ones identified for PCDH11X/Y, allowing an easier and more straightforward analysis of the results. Furthermore, cell and cancer biology studies have been recently carried out specifically for Pcdh10. From an experimental point of view, one of the most relevant reasons of our focus on PCDH10 is the availability of conditional knock-out (KO) mice, generated in our own laboratory (I. Kahr, Ph.D. thesis, Ghent University).

We applied also a proliferation assay based on the measurement of cell metabolic activity. Our preliminary results showed a slightly increased proliferation in cells expressing PCDH11X if compared with non- induced cells: these data agree with the proposed role of protoncogene for PCDH11X but this effect on cell growth needs to be confirmed by repeating the current experiments and by using a complementary proliferation assay. PCDH11X seems to be involved also in adhesion and spreading kinetics in a collagen based assay. Also in this case data are only preliminary since the attempt was to establish in vitro cell based assays for studying specific interactions between PCDHs and the candidate partners identified in our studies. In this experiment PCDH10 did not seem to induce changes in cell behavior if compared with non-induced cells, while PDCH11X apparently induced cells to attach more efficiently to and/or spread more quickly to a collagen-coated surface. Experiments were performed at different time points and this observation is mostly true at the earliest experimental interval. These experiments were carried out mainly with inducible HEK293T FlpIn TRex stable cell lines. The system used can of course influence the results; furthermore, we noticed that in migration and proliferation test the level of expression of the induced protein has a relevant effect. For this reason, efforts to standardize and/or normalize the protein expression will be made in further analyses and experiments.

In conclusion, we established three assays to investigate the roles of specific molecular Pcdh interactions. Further experiments are planned and ongoing, as described. At this stage of the project, the most interesting experiment will involve an attempt to investigate the relation between PCDHs and actin-related proteins. Time-lapse experiments with fluorescent protein variants have shown convincing co-localization of PCDH10 with F-actin filaments, while overlap with microtubules

was not consistent (van Roy Unit, VIB-Ghent University, data not shown). In this context, the discrepancy between long and short isoforms could be indicative for specific roles of these isoforms.

For example, a possible interaction of δPcdhs with the WAVE complex could happen in the context of F-actin re-organization. This could be validated in case the short Pcdh10 isoform was found not to be actin-associated. Moreover, the interaction of the short isoform of Pcdh10 with the WAVE complex would show that the indirect association with NAP1 is not the only form of interaction between the Pcdhs and the WRT, opening interesting perspectives for future investigations.

In vitro experiments have shown that ectopic re-expression of PCDH10 in cell lines where the protein was silenced led to reduced cell migration, cell proliferation and invasive potential [30,31]. In order to reproduce this silencing *in vivo*, and to study the function of Pcdh10 in living organisms, the van Roy Unit (VIB-Ghent University) lab has generated conditional KO mice either for all Pcdh10 isoforms or for its long isoform only (I. Kahr, PhD thesis, Ghent University). In contrast to a previously reported Pcdh10–null mouse, which showed an early lethal phenotype[32], our mice did not show any evident phenotype. Most likely, this is due to differences in the experimental set-up of the study from Hirano, S. *et al.*[3] (S. Hirano, personal communication to F. van Roy) although it could also depend on the different genetic background. Experiments to measure RNA expression levels were conducted in total brain or cerebellum of conditional Pcdh10all KO or Pcdh10long KO mice. QRT-PCR experiments were performed to analyze the expression of δPcdh family members in absence of PCDH10. These experiments showed that the complete loss of Pcdh10 in the early stages of development leads to up-regulations of the other members of the δPcdh family indicating compensation and an overlap in functions between different δPcdhs. Interestingly, the loss of the long isoform of Pcdh10 alone did not induce such an up-regulation of the other δPcdh, but instead triggered a compensatory increase of the short isoform. As the short isoform of PCDH10 lacks the two CM motifs this indirectly indicates that these CMs are not essentially required during early stage development. That compensation event could be the explanation of our intriguing transcriptome analysis conducted on cerebellum of Pcdh10all KO and Pcdh10long KO mice in comparison with wild type (WT) mice. RNA sequencing experiments showed no significant differences in the expression levels of in total KO versus WT mice. In contrast, different expression levels were seen for Pcdh10long KO mice. The analysis of these experiments is still not straightforward since in this particular experiment we could not appreciate up-regulations of the other δPcdhs family members. It would be interesting to perform these transcriptome experiments in total brain and in a larger data set in future follow-up experiments.

The fact that we could not observe the 'spontaneous' development of tumors in Pcdh10 KO mice might indicate that the loss of Pcdh10 alone is not sufficient for cancer formation, but rather it could boost another oncogenic event or a combination of different events. Since mouse Pcdh10 is expressed mainly in brain and because PCDH10 is a documented tumor suppressor candidate in medulloblastoma[33], our lab has generated mouse Pcdh10 ablation models of medulloblastoma. To this end, we have been breeding conditional Pcdh10KO mice with mice undergoing GFAP-Cre-driven loss of p53 (required in many medulloblastoma models) and of retinoblastoma (Rb, which can lead to fast development of medulloblastoma in the external granule layer (EGL) of the cerebellum in combination with the loss of p53). Interestingly, the Pcdh10all KO medulloblastoma model mice (GFAP-Cre // p53 $^{fl/fl\ or\ fl/+}$ // Rb $^{fl/fl\ or\ +/+}$ // Pcdh10-all $^{fl/fl}$) develop medulloblastoma more rapidly and more efficiently than a genetically-related mice with WT Pcdh10.

The list of putative interaction partners proposed from our Array MAPPIT experiments includes many interesting candidates, for which a more thorough study would be relevant. These candidates include Kruppel-like factor 4 (KLF4)(resulted positive for interaction with PCDH9 and PCDH19) and the ADP ribosylation factor-like 13 B (ARL13B, interacting with PCDH11Y, PCDH18 and PCDH19). KLF4 is a member of the Kruppel-like factor (KLF) family of transcription factors and regulates proliferation, differentiation, apoptosis and somatic cell reprogramming.ARL13B, instead, regulates cilia formation and maintenance and is involved in neuronal migration. Future experiments should involve further evaluation of these candidates *in vitro* as well as *in vivo* in mouse tumor models.

Many studies have been performed to date to elucidate the biology of Pcdhs, mostly in non-cancer related experiments using model organisms such as rat and *Xenopus laevis* (reviewed in [1]). δPcdhs play a demonstrated and important role in neurological and neurodegenerative diseases. In particular, *PCDH10* has been proposed to be an autism-spectrum disorder gene[34]. Recently, a study investigated the mechanism which relates PCDH10 to autism, and identified its role as being mostly dependent on elimination of excitatory synapses in the central nervous system[35]. This investigation revealed that Pcdh10 is an important protein mediating the transport of poly-ubiquitinated postsynaptic density protein 95 (PSD-95) to the proteasome. The authors identified a PIR (proteasome interacting region), involved in Pchd10 binding with the proteasome, in the cytoplasmic domain of Pcdh10. The location of this sequence corresponds to the one shared by both short and long isoforms of Pcdh10, i.e. upstream of the CMs. Our array MAPPIT study proposed PSMC1, a proteasome subunit, as Pcdh interaction partner. Therefore, we wanted to investigate this interaction with the aim to demonstrate a role for PSMC1 in the binding of PCDH10 to the

proteasome. Unfortunately, PSMC1 did not coimmunoprecipitate with PCDHs in HEK293Tcells co-transfected with PCDH10 and PSMC1. Proteasomal degradation plays an important role in many processes related to tumor suppressor mechanisms as well. Considering these findings on the interactions involving PSD-95, we expected an increased proteasome-mediated degradation of specific (yet to be identified) target proteins upon Pcdh10 induction, rather than of protocadherins themselves.

Finally, we identified several USPs proteins as potential interaction partners of δPcdhs. USP and other deubiquitinating enzymes (DUBs) can form multi-molecular complexes, with each other and with other proteins. For instance, USP15 interacts with more than 10 partners, including USP46 and USP12. Considering data available so far, the USPs that are able to interact with Pcdhs seem to be involved in signal transduction pathways rather than in other type of cellular events. We performed experiments on ubiquitination and protein degradation for different PCDHs in combination with different USPs in order to investigate the role of these interactions. Our preliminary results showed that the interaction of the USPs with Pcdhs does not influence these mechanisms.

Concluding remarks

Cells control their functions and behavior mostly by the regulation of proteins. Proteins perform in networks but the mechanisms behind such regulations of complex correlations are still not completely elucidated. The "phenotype" of a cell is defined by the state of protein expression which is responsible for the cellular functions in normal development and in pathogenic conditions. Since the human genome was sequenced and many new technologies have been developed, our knowledge concerning the sophisticated cell regulation is exponentially growing. The family of δPCDHs has been shown to be involved in several kinds of cancer and in neurological diseases. Unfortunately, their functions in health are still poorly understood. In this study we identified a list of putative interaction partners, which might further a thorough investigation of the role of such interactions in various cellular mechanisms and signaling pathways. In the van Roy's laboratory Pcdh10 KO mice were generated but we could not observe 'spontaneous' tumor formation in these mice so far, indicating that Pcdh10 defects alone are not sufficient for cancer formation. A Pcdh10all KO medulloblastoma mouse model shows an earlier appearance of the tumors indicating the requirement of a combination of events for cancer development. Identification of interaction partners and the characterization of mouse models for Pcdh-related cancers will be essential to elucidate the role played by PCDHs in signaling pathways and in diverse cellular functions. Elucidating those mechanisms will contribute to cancer research and might in the long run provide new targets for successful anti-tumor therapy development.

References

1. van Roy, F. Beyond E-cadherin: roles of other cadherin superfamily members in cancer. *Nat. Rev. Cancer* **14,** 121–34 (2014).

2. Vanhalst, K., Kools, P., Staes, K., van Roy, F. & Redies, C. delta-Protocadherins: a gene family expressed differentially in the mouse brain. *Cell Mol Life Sci* **62,** 1247–1259 (2005).

3. Nakao, S., Platek, A., Hirano, S. & Takeichi, M. Contact-dependent promotion of cell migration by the OL-protocadherin-Nap1 interaction. *J. Cell Biol.* **182,** 395–410 (2008).

4. Chen, B. *et al.* The WAVE Regulatory Complex Links Diverse Receptors to the Actin Cytoskeleton. *Cell* **156,** 195–207 (2014).

5. Chen, X. & Gumbiner, B. M. Paraxial protocadherin mediates cell sorting and tissue morphogenesis by regulating C-cadherin adhesion activity. *J. Cell Biol.* **174,** 301–313 (2006).

6. Yasuda, S. *et al.* Activity-induced protocadherin arcadlin regulates dendritic spine number by triggering N-cadherin endocytosis via TAO2beta and p38 MAP kinases. *Neuron* **56,** 456–71 (2007).

7. Chen, X., Koh, E., Yoder, M. & Gumbiner, B. M. A protocadherin-cadherin-FLRT3 complex controls cell adhesion and morphogenesis. *PLoS One* **4,** e8411 (2009).

8. Lievens, S. *et al.* Array MAPPIT: high-throughput interactome analysis in mammalian cells. *J Proteome Res* **8,** 877–886 (2009).

9. Hultquist, A. *et al.* Mad 1 Inhibits Cell Growth and Proliferation but Does Not Promote Differentiation or Overall Survival in Human U-937 Monoblasts. *Mol. Cancer Res.* **2,** 464–476 (2004).

10. Hurlin, P. J. & Huang, J. The MAX-interacting transcription factor network. *Semin. Cancer Biol.* **16,** 265–74 (2006).

11. Zhang, H., Fan, S. & Prochownik, E. V. Distinct roles for MAX protein isoforms in proliferation and apoptosis. *J. Biol. Chem.* **272,** 17416–24 (1997).

12. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. The Myc/Max/Mad network and the transcriptional control of cell behavior. *Annu. Rev. Cell Dev. Biol.* **16,** 653–99 (2000).

13. Wahlström, T. & Henriksson, M. Mnt takes control as key regulator of the myc/max/mxd network. *Adv. Cancer Res.* **97,** 61–80 (2007).

14. Blackwood, E. M., Luscher, B. & Eisenman, R. N. Myc and Max associate in vivo. *Genes Dev.* **6,** 71–80 (1992).

15. Barisone, G. A., Yun, J.-S. & Díaz, E. From cerebellar proliferation to tumorigenesis: new insights into the role of Mad3. *Cell Cycle* **7,** 423–7 (2008).

16. Hooker, C. W. & Hurlin, P. J. Of Myc and Mnt. *J. Cell Sci.* **119,** 208–16 (2006).

17. Yun, J.-S., Rust, J. M., Ishimaru, T. & Díaz, E. A novel role of the Mad family member Mad3 in

cerebellar granule neuron precursor proliferation. *Mol. Cell. Biol.* **27,** 8178–89 (2007).

18. Lüscher, B. MAD1 and its life as a MYC antagonist: an update. *Eur. J. Cell Biol.* **91,** 506–14

19. Hurlin, P. J. *et al.* Mad3 and Mad4: novel Max-interacting transcriptional repressors that suppress c-myc dependent transformation and are expressed during neural and epidermal differentiation. *EMBO J.* **14,** 5646–59 (1995).

20. Jung, C.-R. *et al.* Enigma negatively regulates p53 through MDM2 and promotes tumor cell survival in mice. *J. Clin. Invest.* **120,** 4493–506 (2010).

21. Borrello, M. G. *et al.* Differential interaction of Enigma protein with the two RET isoforms. *Biochem. Biophys. Res. Commun.* **296,** 515–22 (2002).

22. Durick, K., Gill, G. N. & Taylor, S. S. Shc and Enigma are both required for mitogenic signaling by Ret/ptc2. *Mol. Cell. Biol.* **18,** 2298–308 (1998).

23. Verset, L., Feys, L., Trépant, A.-L., De Wever, O. & Demetter, P. FHL2: a scaffold protein of carcinogenesis, tumour-stroma interactions and treatment response. *Histol. Histopathol.* **31,** 469–78 (2016).

24. Martin, B. *et al.* The LIM-only protein FHL2 interacts with beta-catenin and promotes differentiation of mouse myoblasts. *J. Cell Biol.* **159,** 113–22 (2002).

25. Gabriel, B. *et al.* Expression of the transcriptional coregulator FHL2 in human breast cancer: a clinicopathologic study. *J. Soc. Gynecol. Investig.* **13,** 69–75 (2006).

26. Ding, L. *et al.* Human four-and-a-half LIM family members suppress tumor cell growth through a TGF-beta-like signaling pathway. *J. Clin. Invest.* **119,** 349–61 (2009).

27. Li, H. Y. *et al.* Protein-protein interaction of FHL3 with FHL2 and visualization of their interaction by green fluorescent proteins (GFP) two-fusion fluorescence resonance energy transfer (FRET). *J. Cell. Biochem.* **80,** 293–303 (2001).

28. Johannessen, M., Møller, S., Hansen, T., Moens, U. & Van Ghelue, M. The multifunctional roles of the four-and-a-half-LIM only protein FHL2. *Cell. Mol. Life Sci.* **63,** 268–84 (2006).

29. Chen, B. *et al.* The WAVE regulatory complex links diverse receptors to the actin cytoskeleton. *Cell* **156,** 195–207 (2014).

30. Ying, J. *et al.* Functional epigenetics identifies a protocadherin PCDH10 as a candidate tumor suppressor for nasopharyngeal, esophageal and multiple other carcinomas with frequent methylation. *Oncogene* **25,** 1070–80 (2006).

31. Yu, J. *et al.* Methylation of protocadherin 10, a novel tumor suppressor, is associated with poor prognosis in patients with gastric cancer. *Gastroenterology* **136,** 640–51.e1 (2009).

32. Uemura, M., Nakao, S., Suzuki, S. T., Takeichi, M. & Hirano, S. OL-Protocadherin is essential for growth of striatal axons and thalamocortical projections. *Nat. Neurosci.* **10,** 1151–9 (2007).

33. Bertrand, K. C. *et al.* PCDH10 is a candidate tumour suppressor gene in medulloblastoma. *Childs. Nerv. Syst.* **27,** 1243–9 (2011).

34. Morrow, E. M. *et al.* Identifying autism loci and genes by tracing recent shared ancestry. *Science* **321,** 218–23 (2008).

35. Tsai, N.-P. *et al.* Multiple autism-linked genes mediate synapse elimination via proteasomal degradation of a synaptic scaffold PSD-95. *Cell* **151,** 1581–94 (2012).

ELEONORA BILLI
Frans Ackermanstraat 27, 9000 Ghent, Belgium
elebilli@msn.com
+32 494 94 04 53
Born: 26/07/1983, Chiusi (SI), Italy
Nationality: Italian

RESEARCH EXPERIENCE

October 2009- September 2016: **Predoctoral research fellow**
in the labs of Professor Frans Van Roy (Molecular Cell Biology Unit, VIB Dept. Inflammation Research centre, Ghent University) and Professor Jan Tavernier (Cytokine receptor Lab, VIB Dept. of Medical Protein Research, Ghent University), Ghent, Belgium.
Topic: *Identification and functional analysis of molecular interaction partners of delta-protocadherin family members with putative roles in cancer.*
Funding: VIB, International PhD Program Fellowship

February 2009-July 2009: **Visiting Research Assistant**
in the lab of Professor Ed Louis, Centre for Genetics and Genomics, Queen's Medical Centre, University of Nottingham, Nottingham, UK.
Topic: *Large scale quantitative genetic analysis of complex traits in yeast.*

October 2007-October 2008: **Research Internship**
in the lab of Professor Carlo Riccardi, Department of Clinical and Experimental Medicine, Department of Pharmacology, Toxicology and Chemotherapy. Perugia University, Perugia, Italy.
Topic: *Glucocorticoid-induced molecular mechanisms.*

January 2006-October 2006: **Research Internship**
in the lab of Professor Carlo Riccardi, Department of Clinical and Experimental Medicine, Section of Pharmacology, Toxicology and Chemotherapy. Perugia University, Perugia, Italy.
Topic: *Glucocorticoid-induced molecular mechanisms.*

EDUCATION

2009-2015**:** **International PhD program** at VIB, Ghent, Belgium.
Ph.D. student in Science: Biochemistry and Biotechnology, Molecular cell biology Unit, VIB Dept. Inflammation Research centre, Ghent University and Cytokine receptor Lab, VIB Dept. of Medical Protein Research, Ghent University.

2006-2008: **Master's degree in Medical Biotechnology,** Faculty of Medicine, University of Perugia,Italy.
Final Degree mark: 110/110 *cum laude*. (First Class with Honours).
Dissertation: "Role of GILZ (*Glucocorticoid-Induced Leucine Zipper*) on the Raf/MEK/ERK pathway and on cell proliferation". Thesis supervisor: Prof. Carlo Riccardi
Award of the "Alessio Trippolini" prize for degree thesis in Molecular Biology.

2002-2006: **Bachelor degree in Biotechnology**, Faculty of Mathematics Physics and Naturals Sciences, University of Perugia, Italy.

Dissertation : “GITR (*glucocorticoid-induced tumour necrosis factor receptor-related gene*) modulates the response to *Candida albicans* infection” .Thesis supervisor: Prof. Graziella Migliorati

PUBLICATIONS

Cubillos FA**, Billi E**, Zörgö E, Parts L, Fargier P, Omholt S, Blomberg A, Warringer J, Louis EJ, Liti G.Assessing the complex architecture of polygenic traits in diverged yeast populations. Mol Ecol. 2011 Apr;20(7):1401-13

NATIONAL AND INTERNATIONAL CONFERENCES

- "VIBes in Biosciences 2010", 2nd International Ph.D. Symposium, October 13-16, 2010, Leuven, Belgium.
- National Cancer Research Institute (NCRI), Cancer Conference, 6-9 November 2010, Liverpool, UK.
- IDIBELL Cancer Conference (ICC) on Metastasis and Angiogenesis, May 26-27, 2011, Barcelona, Spain. Poster presentation: Billi, E., van Roy, F., Tavernier, J.: Identification of molecular interaction partners of delta-protocadherin family members with putative role in cancer.
- 15 years VIB - Closer to the Future, October 6, 2011, Ghent, Belgium.
- 2012 Spring meeting of the BSCDB , “Cell adhesion and cell polarity in development and disease”, April 27-28, 2012 – Ghent, Belgium. Poster presentation: Billi, E., Lieven S., Tavernier, J., van Roy, F.: Identification and functional analysis of molecular interaction partners of delta-protocadherin family members.
- VIBes in Bioscience, 3rd VIB International PhD Symposium, September 5-7 ,2012, Ghent, Belgium.
- Cell Contact & Adhesion Gordon-Merck Research Seminar, June 1-2, 2013, Lucca (Barga) Italy.
- Gordon Research Conference, Cell Contact & Adhesion, June 2-7, 2013, Lucca (Barga), Italy. Poster presentation: Billi, E., Lieven S., Tavernier, J., van Roy, F.: Identification and functional analysis of molecular interaction partners of delta-protocadherin family members with putative role in cancer.
- 30 Years of Plant Biotechnology, November 12th, 2013, Ghent, Belgium.
- 3rd IUAP7/07 1-day meeting DevRepair , December 9, 2013 , in Liège,Belgium. Poster presentation: Billi, E., Lieven S., Tavernier, J., van Roy, F.: Identification and functional analysis of molecular interaction partners of delta-protocadherin family members.
- 4th IUAPVII-07 DevRepair meeting, May 28, 2014, VUB, Brussels-Jette, Belgium
- Yearly VIB seminar 2010, 2012, 2013, 2014, poster presentation, Blankenberge, Belgium.
- VIBes in Bioscience, 4rd VIB International PhD Symposium, September 17-19, 2014, Antwerp, Belgium , poster presentation.

EXTRA

Student supervision for Erasmus and bachelor projects
Organization Committee of PhD student day at the VIB seminar, Blankenberge, April 2012
Organization Committee of VIBes in Bioscience, 3rd VIB International PhD Symposium, September 5-7 ,2012, Ghent, Belgium.
Collaborator for the organization of Spring meeting of the BSCDB , “Cell adhesion and cell polarity in development and disease”, April 27-28, 2012 – Ghent, Belgium.
Participation at the CEITEC Scientific Retreat, October 3-4, 2013, Brno, Czech Republic, presentation of the VIB International PhD program.

Acknowledgements

One of the first things I learned when I started my PhD was that experiments done at the weekend do not work, but despite this everyone keeps doing it. The truth is that the chances that an experiment will fail are much higher than the possibility of success. Telling yourself that it's the weekend's fault is just looking for an excuse to give up. So basically what this time taught me is the importance of steadfastness in whatever direction you choose, that tiredness is a state of mind, Sunday is just a name, that satisfaction occurs after perseverance. And here I am: this thesis is the results of years of personal learning and growing for which I am deeply and truly thankful.

I want to thank my PhD examination committee for their precision and precious comments to improve the quality of the thesis Prof. Dr. Johan Grooten, Prof. Dr. Geert Berx, Prof. Dr. Jean Christophe Marine, Prof. Dr. Marleen van Troys, Dr. Irma Lemmens, Dr. Karl Vandepoele.

I would like to thank the Vlaams Instituut voor Biotechnologie (VIB), first because being part of the VIB International PhD programme has been a source of pride and satisfaction and especially I would like to thank Marijke, Mark and Lieve who shepherded me through this experience. It was a pleasure for me to present the programme at the CEITEC institute in Brno, because my experience as a VIB PhD student was valuable and of great satisfaction. Second, I am thankful for the possibility to work in such an international environment, full of different skills, tools, opportunities and easy connections between groups and institutes which showed me how collaboration, communication and sharing is the key to perform high level science.

This work would not have been possible without my promoters, who designed the project and welcomed me into their groups, who supported me and the project during those years with meetings, discussions, criticism and advice in a high-level scientific environment which helped me to grow as a scientist and as a person: Prof. Dr. Frans van Roy, who I want to thank especially for teaching me the importance of knowledge and literature, and commitment to a research project — I really appreciated all the support I received during the writing of my thesis despite the (disappointing) results — and Prof. Dr. Jan Tavernier who I want to thank for the scientific support and advice and for giving me the opportunity to be part of a challenging and beautiful place like his lab.

Special thanks go to all the members of the FVR unit for scientific and moral support. To Vanessa, my unofficial tutor who committed her time and knowledge to teach me most of the things I learned during the project, for her unconditioned friendship and support, for her constant will to keep an eye on me, for always being next to me. Vane, I will never thank you enough. To Irene, my "protocadherin buddy" as she says, for sharing with me the learning process, protocols, desk, results, frustrations, success and laughter. To Mara, my little sweet Mara, for all the energy and joy she brought in the lab but above all for all the love she always showed me. To Sahin, an example of commitment and passion to this job, for the conceding me the two days a week I was allowed to complain, for the support and energy he gave me in the never-ending writing/correcting/defending

process. To Katrien, for being the person who I could count on for the organization of everything. The first memory of her I have dates back to my first Christmas abroad, when I thought I could not join my family because of the snow and she invited me to join her family. I will never forget it. To Ellen, for all the positive and encouraging words she gave me over the years. To Koen for sharing the years of PhD frustrations, for the scientific talks and for the jokes which were always needed to interrupt my typically negative stream of thoughts. To Vidha, for her sweet heart, for sharing with me the richness of her world so far from mine. To Evi, for the fresh energy she brought to the lab. I would also like to thank those who were part of the group when I arrived for their warm welcome, for their smiles, for making the group so diverse and rich like Tim, Tanya, Huiyu, Philippe, Ilse, Jolanda and everyone who spent some time with us, such as Tom and Martina. I want to thank my friend Elisabeth especially, for all the lovely and special time we spent together.

I also really want to thank all the members of the Cytokine Receptor Lab who welcomed me every time I came back to learn something new, for advice, for expertise. It is a beautiful group of people, when it comes to knowledge and willingness to share. I am truly grateful for the opportunity to be part of this group and for the really nice time I had working with all of them. I would especially like to thank Sam, for being involved in supporting my project and helping with information and comments every time I was in need; and, Dieter for taking care of my precious MAPPIT arrays and helping me any time I had to face the umpteenth MAPPIT experiment. To Irma, for sharing her experience of research life with me, for listening to my PhD-related fears, for always showing me that is not so dark as it looks, for always being encouraging and also for being part of my examination committee. To Laura, my friend Laura, the first person I meet when Belgium was still an undefined place, far from becoming home. Thank you, Laura, for the walk around the institute before day 0, for the encouraging chat and for welcoming me to the lab environment, to your house and to life in Ghent, for all the times, the chats, the moments we shared, for all the memories which I will always carry with me. To Ioanna, though I don't even know where to start... Thank you, Greca, for being here with me in this strange time, for your scientific advice, for your special friendship, your time, your listening ear, your dinners, your hugs, your delicate respect for someone else's feelings and thoughts. I want to thank the "girls": Delphine, Jennyfer, Leentje, for being with me when my world turned upside down, for showing me that actually the only direction which matters is our own. Thank you also, girls, for the support you gave me with my project, teaching me your expertise, sharing your concern, for your always-encouraging words. To Anne Sophie, because of her willingness to help me with reagents so I could save time in my weird schedule and for the support I always found in her understanding smile. And the "boys", the PhD gang: Kevin and Darek, for the nice time we had together outside the lab and for your scientific support, always precious; Raffaele for being part of this journey in many different ways, in particular for the fact that I could count on you to understand me; Krzysztof, for the nice chats and your positive attitude, Tim for your contagious energy. To Karolien, because I deeply admire her and because few years ago she send a mail which is still on my wall ending with "it's not like baking a cake and putting the oven on too long, then at least you know why your cake is black. Science is not like that, eh. Keep up the confidence." Thanks to Anje, for every time she asked about my work, for her contagious energy, her encouraging words, her positive and strong attitude. I would like to sincerely thank Lennart, because

I always felt I could count on him; Josè for his kindness and never-ending support; Nele for the lovely company and smiles; Jos, for the face he made when my pile of transfected 96 well plates dropped on the floor on a Sunday morning and for being such a pleasant company in the lab; Dominique and Annick for their expertise and technical support and their positive attitude; Julie, Sophie, Elien, Steffi, Lode, Virginie, Laurence, Tim, Nadia, Felien, Thomas, Isabel and Viola: this special group wouldn't be the same without your contribution. It was a real pleasure to work with all of you.

Special thanks go to Chiarina who crossed my path in this at the time unfamiliar institute, and since then walked by me every step of those years. This time would have never been the same without you, my friend, my Person. Thank you for showing me a beautiful way of approaching life: open, curious, honest, altruistic. Thank you for being part of my life: you are the proof that certain people were meant to meet. I want to thank the group of initially colleagues, now lovely friends, who guided me on my arrival, slowing down their English, showing me the way, especially my beautiful and precious friends Naira and Stephanie, because we have got a feeling, and we were right.

I want to thank all the people from the two departments who helped me over the years making PhD life easier, especially Marita, Myriam and Chantal, but also Rik, Wim, Veronique, Dirk, Johan and Serge. A special thought to the PhD writing gang of the library (Sahin, Laura, Alice): thank you for the precious company and priceless help: keep going, now it's your turn!

This time has been intense, my PhD was not only a project, but a new life: it was of inestimable value to be surrounded by special people outside the labs that's why I still would like to thank every person who shared this journey with me in any way, I want to thank all of you deeply for making me feel home. It was a special gift, and there are no words which can thank you enough. A special thank goes to Pablo and Ludo, for giving me a place to stay, a place to complain, a place to laugh and for introducing me to their special families and fabulous friends, the musician gang, my beloved Floozies, which become my world here and made Ghent my home away from home. To Pieter and Bieke, for making me the proud Godmother of my Robbino. To Alessia, a loud thank you, able to reach her on the other side of the world, for everything she knows. I cannot conclude without mentioning Paola for every precious hug, confession and moment we shared, Nele, because certain evenings would have broken me without her, Niels and Petra, for constantly caring about me, Adriana, because she is an amazing person and for all the incredible help she gave me also in these testing times, Jasper and Yasmina, because their company has always been a reason to smile, and a warm thanks to all the other friends which belong to this beautiful group which made this time so pleasant and memorable. When you make a choice you leave something, by definition, but I want to thank my "old" friends who did not forget about me and that never really let me go despite life and distance, showing that loving someone is not a choice, especially Chiara who never left me alone since that phone call after the interview, for the time she shared with me in Ghent and in life; my beloved Francesca, my "North"; Andrea, Leonardo, Alessandra for far too many reasons. I also would like to mention and to thank Prof. Carlo Riccardi for being my scientific mentor and to Don Antonio, my spiritual guide.

I would like to thank everyone who kept calling when I was far or distant, who invited me to their special days, came to water my plants when I forgot what I cared about, brought me food in the library or when I had no time, didn't leave me alone with boxes to move, mice to find, broken batteries, success to share; who called to check if I was safe, wondering if I was happy. I want to thank who asked about my work, about my thoughts, about my wonders; who shared being part of this city with me; who showed me places and directions.

And I want to give thanks with all my heart to whom made me feel proud of who I am, who was never far away, who proved to me that time and distances do not matter and that it is true that love is the answer, brightening my northern sky. Thank you, for the answer and for listening to the words that hadn't been invented yet.

And then my family. I nonni, my precious, precious nonni, for understanding and discounting the distance. La mamma e il babbo, because the day I left they pretended they did not have tears and because nothing would have been possible without them. Thanks for telling me I could do it, for being an example of constancy and responsibility, for loving me.

Finally I want to thank my brother Alessandro, with everything I have, for always standing by my side, for the words he said and for those unspoken and for loving me no matter what ("let them; but first we pass"). I miss you every day.

E la mia famiglia. Grazie ai miei preziosissimi nonni per aver capito e perdonato la distanza, e per averla cancellata. E poi grazie, ma' e ba', perchè quando sono partita avete sorriso e nascosto le lacrime e perchè non sarebbe stato possibile niente senza il vostro appoggio, la vostra spinta, la vostra fiducia e stima. Grazie per non aver mai avuto dubbi, per essere un esempio di costanza e responsabilità, per tutto il vostro amore e per aver sempre fatto il tifo per me.

E poi, grazie a te, Ale, per essere stato sempre al mio fianco senza dubbio, per ogni parola che hai detto, per quelle che hai nascosto lontane da me e per volermi quel bene lì, a prescindere da tutto ("facciano pure; ma prima noi passiamo"). Mi manchi sempre.

I tried to hold back my typical outpouring of emotions, I have left a lot of important things unsaid, but today I am grateful for what I have achieved and for who I have become and this is also because of you. Thank you all for being part of my (PhD) life. I will never forget it.

"I took a deep breath and listened to the old brag of my heart.
I am, I am, I am."(Sylvia Plath, The Bell Jar)

Ghent, 22.09.2016

Eleonora

Cover :

Picture from Sakai *et al*., 2011. “Protein interactome reveals converging molecular pathways among autism disorders”.

(The extended ASD protein interaction network. Nodes and lines indicate the 11 syndromic ASD proteins (red), the 23 ASD-associated proteins used as bait in Y2H (pink), new binding partners identified in the study (purple), known binding partners (orange), new interactions (green), and previously known interactions (light blue).)

Graphical design : www.leendewver.com

[From below comes the noise of a door slamming.]

(Henrik Ibsen, A Doll's House)